

10/600-862

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 4, 2005, 11:07:54 ; Search time 15263 Seconds
(without alignments)
11752.684 Million cell updates/sec

Title: US-10-600-862A-6
Perfect score: 3702
Sequence: 1 cttaccgcatgagggtcc.....tcagctggtgtgactcgag 3702

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2647.8	71.5	3624	6	AX840238 Sequence
2	2647.8	71.5	3624	9	Y00839 H. sapiens G
3	2647.8	71.5	3846	6	CQ882052 Sequence
4	2646.2	71.5	3625	6	CQ723359 Sequence
5	2639.8	71.3	3441	9	M34424 Human acid
6	2635	71.2	3687	9	BC040431 Homo sapi
7	2585.8	69.8	3626	11	BV177050 sqnm92407
8	2033.2	54.9	2814	4	AF171665 Bos tauru
9	1903	51.4	3408	10	BC061753 Rattus no
10	1899.8	51.3	3504	10	BC010210 Mus muscu
11	1881.4	50.8	3364	6	AX305670 Sequence
12	1881.4	50.8	3364	10	MMU49351 Mus muscu
13	1332	36.0	3112	5	AB006754 Coturnix
14	994.4	26.9	3548	5	AB000967 Coturnix
15	957	25.9	1493	6	CQ776406 Sequence
16	957	25.9	1493	6	AX052732 Sequence
17	957	25.9	1493	6	AX557300 Sequence
18	957	25.9	1493	9	HUMA2MGRAP Human alpha
19	955.4	25.8	1519	9	HSM808130 Homo sapi

20	953.8	25.8	1499	6	BD134425	Human nuc
21	953.8	25.8	1499	6	AX017254	Sequence
22	953.8	25.8	1499	6	AX524958	Sequence
23	895.4	24.2	1485	6	CQ718547	Sequence
24	665.6	18.0	1419	10	BC057979	Mus muscu
25	665.6	18.0	1451	10	BC032170	Mus muscu
26	665.6	18.0	1478	10	867967	heparin bin
27	665.6	18.0	2902	6	CQ777321	Sequence
28	665.6	18.0	2902	6	AX305444	Sequence
29	665.6	18.0	2902	10	MUSHBP44	Mus muscu
30	665.6	18.0	2913	10	BC046641	Mus muscu
31	665.6	18.0	1373	10	BC082020	Rattus no
32	665.4	17.9	1452	10	AJ639614	Mus muscu
33	663.8	17.9	2768	10	RN45KDA	Mus muscu
34	616.6	16.7	2753	10	BC059887	Mus muscu
35	581.6	15.7	1260	10	RATGP330	Rat Heymann
36	509	13.7	1624	6	BD134483	Human nuc
37	509	13.7	1624	6	AX017383	Sequence
38	509	13.7	1624	6	AX525087	Sequence
39	465.8	12.6	1497	5	GD4J1912	Gallus do
40	424.4	11.5	969	11	BV179107	sqnm10389
41	415.4	11.2	522	11	BV168003	sqnm7017
42	388.6	10.5	2873	8	AK105449	Oryza sat
43	387.2	10.5	1853	5	BC054293	Xenopus l
44	382.8	10.3	2804	8	AK121588	Oryza sat
45	371.2	10.0	6513	9	AF016833	Homo sapi

ALIGNMENTS

RESULT 1
AX840238
LOCUS AX840238 3624 bp DNA linear PAT 16-DEC-2003
DEFINITION Sequence 12 from Patent WO03073839.
ACCESSION AX840238
VERSION AX840238.1 GI:39978638
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Fogher, C. and Reggi, S.
TITLE Expression of lysosomal enzymes in plant seeds
JOURNAL Patent: WO 03073839-A 12 12-SEP-2003;
FEATURES
Location/Qualifiers
source
1..3624
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
sig_peptide 220..426
CDS <427..3078
/note="unnamed protein product"
/codon_start=1
/protein_id="CAE85448.1"
/db_xref="GI:39978639"
/translation="AHPGRPRAPVPTQCDVPPNSRFDCAPKAITQBCARGCCYTPA
KQGLQAGNQWCFPPSPYSKYKLENLSSEMGYTATLTRTPTTFEPKDLILRLDVL
MMETENRLHFTIKDPANRYEVEPLETPRVHSRAPSPLYSVFSESEFFGVIVHRLDGR
VLLNTTVAFLPFADQFLQSLTSPQYITGLAEHLSPMLSTSWITLWNRLDPTP
BKLYGSHHPFYLALEDGSAHGVLNNSAMDVLPQSPALSWRSTGGILDVYFLGP
EPKSVQOYLDVVGYPFMPYMGVGLFHLRCWGYSSATIRQVVENNTRAHFPLDQWN
DLIDWNSRDET FNKGDGDRFPAMQELHQGRRYMMI VDPATISSGSPAGSVRYPDEG
LRGFVITNETQPLIGKWPSTAFPDFTNTALAMWEDMAEFHQVDFDQWIDM
NPSFPIRGSEGCNNELNPPYFVGVGGLQATATCASHQFSLTHNLNLVGL
TEAISHALVARGTRPFVIRSFAGRYAGHWTDVWSWSEGLASVSPILQFN
LIGVLVGVADVCQGLTSEELCVMTQLGAFYPMRNHNSLSLSPQYSPSEFAQQ
AMRKALTLRYALLPHLYTLFHOAHVAGETVARPLFLEFPKDSSTWTVDQLLMGEALL
ITPVLQAGKAEVTGYPLGTVDLOTVPTEALGSLPPPPAPAPREPAIHSEGWVTLPA
PLDTTNVHLRAGYIIPLOGGLTTTESRQQPALAVLTKGGEARGELFDDGESLEVL
LERGAYTQVIFLARNTINVELRVSTSEGAGLQQLQKVTVLGVATAPQQVLSNGVPVSN
FTYSPDTKVLIDICVSLLMGEQLVSNWC"

[illegible]

Qy	3136	TTCCCTCTGGGCACATGGTACGACCTGCAGACGGTCCCAATAGAGCCCTTGGCAGCCTC	3195
Db	2518	TTCCCTCTGGGCACATGGTACGACCTGCAGACGGTCCCAATAGAGCCCTTGGCAGCCTC	2577
Qy	3196	CCACCCCACTGACCTCCCGTAGCCAGCCATCCACGAGGGGCACTGGGTGACG	3255
Db	2578	CCACCCCACTGACCTCCCGTAGCCAGCCATCCACGAGGGGCACTGGGTGACG	2637
Qy	3256	CTGCGCGCCCTCGACCAACACCTCCACCTCCGGGCTGGGTACATCATCCCCCTG	3315
Db	2638	CTGCGCGCCCTCGACCAACACCTCCACCTCCGGGCTGGGTACATCATCCCCCTG	2697
Qy	3316	CAGGGCCCTGGCTCTCAACACACAGAGTCCCGCAGAGCCCATGCGCTGTGGCC	3375
Db	2698	CAGGGCCCTGGCTCTCAACACACAGAGTCCCGCAGAGCCCATGCGCTGTGGCC	2757
Qy	3376	CTAACCAAGGTGGAGAGCGCCGAGGGAGCTTCTGGGACGATGGAGAGCCCTGGAA	3435
Db	2758	CTAACCAAGGTGGAGAGCGCCGAGGGAGCTTCTGGGACGATGGAGAGCCCTGGAA	2817
Qy	3436	GTGCTGGAGCGAGGGCCTACACACAGGTCTCTTCTGGCCAGGGAATAACACGATCGTG	3495
Db	2818	GTGCTGGAGCGAGGGCCTACACACAGGTCTCTTCTGGCCAGGGAATAACACGATCGTG	2877
Qy	3496	AATGAGCTGGTACGTGTGACCAAGTGAAGGAGCTGGCCTGCAGCTGCAGAGGTACTGTC	3555
Db	2878	AATGAGCTGGTACGTGTGACCAAGTGAAGGAGCTGGCCTGCAGCTGCAGAGGTACTGTC	2937
Qy	3556	CTGGGCGTGGCCACGCGCCCGCCAGAGTCTCTTCCAAAGGTGTCCTGTCTCCAACTTC	3615
Db	2938	CTGGGCGTGGCCACGCGCGCCCGCCAGAGTCTCTTCCAAAGGTGTCCTGTCTCCAACTTC	2997
Qy	3616	ACCTACAGCCCGACACCAAGGTCCTGGACATCTGTCTCGCTGTTGATGGAGAGCAG	3675
Db	2998	ACCTACAGCCCGACACCAAGGTCCTGGACATCTGTCTCGCTGTTGATGGAGAGCAG	3057
Qy	3676	TTTCTCGTCAGCTGGTGT 3694	
Db	3058	TTTCTCGTCAGCTGGTGT 3076	
RESULT 2			
LOCUS	HSGAA	3624 bp	mRNA
DEFINITION	H.sapiens GAA mRNA for lysosomal alpha-glucosidase (acid maltase).		
ACCESSION	Y00839		
VERSION	Y00839.1 GI:31607		
KEYWORDS	alpha-glucosidase; amylase; glycoprotein; lysosomal enzyme.		
SOURCE	Homo sapiens (human)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1		
AUTHORS	Hoefsloot, L.H., Hoogeveen-Westerveld, M., Kroos, M.A., van Beemten, J., Reuser, A.J., and Oostra, B.A.		
TITLE	Primary structure and processing of lysosomal alpha-glucosidase; homology with the intestinal sucrase-isomaltase complex		
JOURNAL	EMBO J 7 (6), 1697-1704 (1988)		
MEDLINE	8905058		
PUBMED	3049072		
REMARK	(revised by (31))		
REFERENCE	2		
AUTHORS	Reuser, A.J.J.		
TITLE	Direct Submission		
JOURNAL	Submitted (24-JUN-1988)		
REMARK	(revised by (31))		
REFERENCE	3 (bases 1 to 3624)		
AUTHORS	Reuser, A.J.J.		
TITLE	Direct Submission		
JOURNAL	Submitted (08-JUN-1990) Reuser A.J.J., Department of Cell Biology and Genetics Erasmus University, P.O. Box 1738, 300 DR Rotterdam, Netherlands		
COMMENT	Data kindly reviewed (08-JUN-1990) by Reuser A.		

FEATURES	Source
Location/Qualifiers	
1..3624	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/issue_type="placenta and testis"
	/clone_lib="lambda gt11 (testis) and lambda gt11 (VII-75-1; placenta A1)"
mRNA	<1..3624
	/product="alpha-proglucosidase mRNA"
gene	220..3078
	/gene="GAA"
CDS	220..3078
	/gene="GAA"
	/EC_number="3.2.1.3"
	/codon_start=1
	/product="glucan 1, 4-alpha-glucosidase"
	/protein_id="CAA68763.1"
	/db_xref="GI:31608"
	/db_xref="GOA:P10253"
	/db_xref="UniProt/Swiss-Prot:P10253"
	/translation="MGVRHPPCSHLLAVCALVSLATAALLGHILLHDLFLVPRELSSPVLEETHPAHQGASRGRDQAHPGRPEAVPTQCDVPNSRFDCAPDKAITOEQCEARCCCIYPAKQGLQAGMGPWFPPSPSYKLENLSSEMGCYATLITPTTFPKDILTLRDVMETENRLHFTIKDPANRRYEVPLETRVSRAPSPLSVFSBEPFVIVHRQDGRVLLNTTVAFLPFAFDQLQSLSPSYITGLAHLSPMLSTSWTRI TLNRRDLAPTGANLYGSHFYLALEDGSAHGVLFLNSNAMDVLQSPALSWRSTG GILDVYIFLGPEPKSVQYLDVGVYPMPPYMGFLGFLCRWGYSTAITROVENVNT RAHFPLDVQMDLDYMDSRDPTFNKDGFRDEPMVQELHOGGRYMMIVDAISSG FAGSTRPDEGLURGVFITNETGQLIKGWFGSTAFDFTNPTALLAWHEDVMAEPHD QVFPDGMIDMNEPNSFIRSGEDGCPNELENPPYVGVGGTLOQAATCASHHQFLS THYNLHNYLGTETAIASHRALVKARTRPFVISRSTFAGHRYAGHTGDVSSWEQL ASSVEIPQLNLLGVLGADVCGFLGNTSELCVWTQLGAFYFPMRHNLSLLSPQ BPSYSEPAQAMRKALTRYALPHLYTLPHQAVAGETVARPLFLEPKDSSTWTV DHQLWGSAALIITPVLAGKAEVTCYFPLGTWYDLQTVPIEALGSLPPPAAPREPAI HSEGWTLDPAPLDITNVHLRAGYIIPLOGPLTTTSESRQOPMALAVALTKGEARGE LFDGDESLEVLERGAYTQVIFLARNTIIVNELVRTSEGAGLQKQVTLGVATAPQ QVLSNGVPVSNFTYSPDTKVLDCVSLLMGEQFLVSWC"
sig_peptide	220..426
mat_peptide	/gene="GAA"
	427..3075
	/gene="GAA"
	/product="glucan 1, 4-alpha-glucosidase"
	/EC_number="3.2.1.3"
misc_feature	427..828
	/gene="GAA"
	/note="propeice of 70 kD alpha-glucosidase (AA 1 - 134)"
misc_feature	427..582
	/gene="GAA"
	/note="propeice of 76 kD alpha-glucosidase (AA 1 - 52)"
misc_feature	<583..3075
	/gene="GAA"
	/note="76 kD alpha-glucosidase (AA 54 - <883)"
misc_feature	637..639
	/gene="GAA"
	/note="pot. N-linked glycosylation site"
CDS	<829..3078
	/gene="GAA"
	/note="no start codon"
	/codon_start=1
	/product="70 kD alpha-glucosidase"
	/protein_id="CAA68764.1"
	/db_xref="GI:31609"
	/db_xref="GOA:P10253"
	/db_xref="UniProt/Swiss-Prot:P10253"
	/translation="APSPHYSVEFSEBEPFGLVHRQLDGRVLLNTTVAFLPFADQFLQ LSTSPSQSYITGLAHLSPMLSTSWTRITLNRDLAPTPGANLYGSHFYLALEDGG SAHGVLFLNSNAMDVLQSPALSWRSTGILDVYIFLGPEKSVQYLDVGVYPM RDPFMVQELHOGGRYMMIVDPAJSSGCPAGSYRYPDEGLRRGVFITNETQPLIGK VHPGSTAPDFTNPTALAWEDMVAEFHDQVFPDGMWIDMNEPSPFIRSGEDGCPNE LENPPYVGVGGTIOQAATCASHHQFLSTHYNLNLGLTETAIASHRALVKARTRP FVISRSTFAGHRYAGHTGDVSSWEQLEASSVPEILQFNLLGLVPLGVADVCGFLGNT

1360	GCACACGGGGTGTTCCTGCTTAAACAGCAATGCACATGATGTGGTCTCTGCAGCCGAGCCCT	1419
1816	GCCTTTAGCTGGAGGTCCGACAGGTGGGATCCTGGATGCTTACATCTTTCTCTGGGCCACAGAG	1875
1420	GCCCTTAGCTGGAGGTCCGACAGGTGGGATCCTGGATGCTTACATCTTTCTCTGGGCCACAGAG	1479
1876	CCCAAGAGCGTGTGCAGCAGTACCTTGGACGTTTGTGGGATACCGTTTTCATGCGCCCATAC	1935
1480	CCCAAGAGCGTGTGCAGCAGTACCTTGGACGTTTGTGGGATACCGTTTTCATGCGCCCATAC	1539
1936	TGGGGCTCTGGGCTTTCACACTGTGCGCTGGGGGTACTCTCCACCGCTATATCAACCOCGCAG	1995
1540	TGGGGCTCTGGGCTTTCACACTGTGCGCTGGGGGTACTCTCCACCGCTATATCAACCOCGCAG	1599
1996	GTGGTGGAGAACATGACACAGGGCCCACTTCCCCTTGGACGTCTCMAATGGAAACGACTGGAC	2055
1600	GTGGTGGAGAACATGACACAGGGCCCACTTCCCCTTGGACGTCTCMAATGGAAACGACTGGAC	1659
2056	TACATGACTTCCCGAGGGGACTTCACAGTTCMAAACAAGGATGGCTTCCGGGACTTCCCGGGCC	2115
1660	TACATGACTTCCCGAGGGGACTTCACAGTTCMAAACAAGGATGGCTTCCGGGACTTCCCGGGCC	1719
2116	ATGGTGCAGAGCTGCACCAAGGGCGGCCGCTTACATGATGATCGTGGATCTGTGCATC	2175
1720	ATGGTGCAGAGCTGCACCAAGGGCGGCCGCTTACATGATGATCGTGGATCTGTGCATC	1779
2176	AGCAGCTCTGGGCCCTGTCCGGAGGACTCAGAGCCCTTACGACGAGGGTCTGCCGAGGGGGGTT	2235
1780	AGCAGCTCTGGGCCCTGTCCGGAGGACTCAGAGCCCTTACGACGAGGGTCTGCCGAGGGGGGTT	1839
2236	TTCATCAACAGAGACGGCCAGCCGCTGATTGGGAAGGTATGGCCCGGGTCCACTGCC	2295
1840	TTCATCAACAGAGACGGCCAGCCGCTGATTGGGAAGGTATGGCCCGGGTCCACTGCC	1899
2296	TTCCCCGACTTACCAACCCCAACGCCCTGGCTGGGAGGACATGCTGGCTGAGTTCT	2355
1900	TTCCCCGACTTACCAACCCCAACGCCCTGGCTGGGAGGACATGCTGGCTGAGTTCT	1959
2356	CATGACCAGGTGCCCTTTCGACGGCTTCTGTGATTGACATGAACAGAGCCCTTCCAACTTCATC	2415
1960	CATGACCAGGTGCCCTTTCGACGGCAATGTGATGACATGAACAGAGCCCTTCCAACTTCATC	2019
2416	AGAGGTCTTGAGGACGGCTGCCCAACAATGAGCTGGAGAACCCACCTAGTGGCTGGG	2475
2020	AGAGGTCTTGAGGACGGCTGCCCAACAATGAGCTGGAGAACCCACCTAGTGGCTGGG	2079
2476	GTGGTTGGGGGACCCCTCCAGGGGGCCACCATCTGTGCCCTCCAGACCAACAGTTTCTCTCC	2535
2080	GTGGTTGGGGGACCCCTCCAGGGGGCCACCATCTGTGCCCTCCAGACCAACAGTTTCTCTCC	2139
2536	ACACATPACAACTTGCAAAACCTTCTACGGCCTTGACCGAAGCCATCGCCTCCCAAGGGCG	2595
2140	ACACATPACAACTTGCAAAACCTTCTACGGCCTTGACCGAAGCCATCGCCTCCCAAGGGCG	2199
2596	CTGGTGAAGCTCGGGGACACGCCCATTTTGATCTCCCGCTCGACCTTTGTGGCCAC	2655
2200	CTGGTGAAGCTCGGGGACACGCCCATTTTGATCTCCCGCTCGACCTTTGTGGCCAC	2259
2656	GGCCGATACCGCCGCCACTGGAACGGGGACGTGTGAGCTCTCTGGGACGAGCTCGCCTCC	2715
2260	GGCCGATACCGCCGCCACTGGAACGGGGACGTGTGAGCTCTCTGGGACGAGCTCGCCTCC	2319
2716	TCCGTGCCAGAAATCTTGCAAGTTTAAACCTGTGGGGGTGCTCTGTGTGGGGGCGACGCTC	2775
2320	TCCGTGCCAGAAATCTTGCAAGTTTAAACCTGTGGGGGTGCTCTGTGTGGGGGCGACGCTC	2379
2776	TGGGGCTTCTTGGGCAACCTCAGAGGAGCTGTGTGTGGCTGTGACCCACAGCTTGGGGGCC	2835
2380	TGGGGCTTCTTGGGCAACCTCAGAGGAGCTGTGTGTGGCTGTGACCCACAGCTTGGGGGCC	2439
2836	TTCTACCCCTTCAATGCGGAAACCAACACAGCCTGTCTAGTTCTGCCCCCAGGAGCCGTACAGC	2895
2440	TTCTACCCCTTCAATGCGGAAACCAACACAGCCTGTCTAGTTCTGCCCCCAGGAGCCGTACAGC	2499

Qy	2896	TTAGCGAGCCGGCCCA	CGAGCCATGAGGAAGCCCTC	ACCTCGCGTACGCACTCCTC	2955		
Db	2500	TTAGCGAGCCGGCCCA	CGAGCCATGAGGAAGCCCTC	ACCTCGCGTACGCACTCCTC	2559		
Qy	2956	CCCACCTCTACACACT	GTTCACACAGGCCACAGT	CGCGGGGAGACCGTGGCCCGGCCC	3015		
Db	2560	CCCACCTCTACACACT	GTTCACACAGGCCACAGT	CGCGGGGAGACCGTGGCCCGGCCC	2619		
Qy	3016	CTCTTCCTGGAGTTCC	CAAGSACTCTAGCACCTG	SACCTGACACCAAGCTCTGCTG	3075		
Db	2620	CTCTTCCTGGAGTTCC	CAAGSACTCTAGCACCTG	SACCTGACACCAAGCTCTGCTG	2679		
Qy	3076	GGGAGGCCCTGCTCAT	CCCCAGTGTCTCCAGGCG	GGGAGGCCGGAAGTACTTGGCTAC	3135		
Db	2680	GGGAGGCCCTGCTCAT	CCCCAGTGTCTCCAGGCG	GGGAGGCCGGAAGTACTTGGCTAC	2739		
Qy	3136	TTCCCTTTGGGCACAT	GTGTACGACCTGACAGCG	TGCAATAGAGGCCCTTGGCAGCCTC	3195		
Db	2740	TTCCCTTTGGGCACAT	GTGTACGACCTGACAGCG	TGCAATAGAGGCCCTTGGCAGCCTC	2799		
Qy	3196	CCACCCCACTGCGAGCT	CCCGCTGACGAGCCATCCA	CAGCAGGAGGCCAGTGGGTGACG	3255		
Db	2800	CCACCCCACTGCGAGCT	CCCGCTGACGAGCCATCCA	CAGCAGGAGGCCAGTGGGTGACG	2859		
Qy	3256	CTGCGGCCCCCTGGACA	CCATCAACGCTCCACTCCG	GCTGGGTACATCATCCCCCTG	3315		
Db	2860	CTGCGGCCCCCTGGACA	CCATCAACGCTCCACTCCG	GCTGGGTACATCATCCCCCTG	2919		
Qy	3316	CAGGGCCCTGGCTCA	CAACACAGAGTCCCGCAG	CAGCCCATGGCCCTGGCTGTGGCC	3375		
Db	2920	CAGGGCCCTGGCTCA	CAACACAGAGTCCCGCAG	CAGCCCATGGCCCTGGCTGTGGCC	2979		
Qy	3376	CTAACCAAGGTCGAGAG	CGCCGAGGGAGCTGTCTTG	GGACCATGAGAGAGCCCTGGAA	3435		
Db	2980	CTGACCAAGGTCGAGAG	CGCCGAGGGAGCTGTCTTG	GGACCATGAGAGAGCCCTGGAA	3039		
Qy	3436	GTGCTGGAGCGAGGGC	CTACACACAGGTCATCTT	CTGGCCAGGAATTAACGATCGTG	3495		
Db	3040	GTGCTGGAGCGAGGGC	CTACACACAGGTCATCTT	CTGGCCAGGAATTAACGATCGTG	3099		
Qy	3496	AATGAGCTGTGACGTG	TGACCATGAGGAGCTGGCT	TCGAGCTGCAGAGGTGACTGTC	3555		
Db	3100	AATGAGCTGTGACGTG	TGACCATGAGGAGCTGGCT	TCGAGCTGCAGAGGTGACTGTC	3159		
Qy	3556	CTGGGCTGGCCACGGC	CGCCCGCAGCAGGTCTCT	CTCAAACGGTGTCCCTGTCTCCAACTTC	3615		
Db	3160	CTGGGCTGGCCACGGC	CGCCCGCAGCAGGTCTCT	CTCAAACGGTGTCCCTGTCTCCAACTTC	3219		
Qy	3616	ACCTACGCCCCGACACA	AGAGTCTTGACATCTGTG	TCCTCGCTGTGATGGAGAGCAG	3675		
Db	3220	ACCTACGCCCCGACACA	AGAGTCTTGACATCTGTG	TCCTCGCTGTGATGGAGAGCAG	3279		
Qy	3676	TTTCTCGTCAGCTGGT	TTT 3694				
Db	3280	TTTCTCGTCAGCTGGT	TTT 3298				
RESULT 4							
LOCUS		CQ723359		3625 bp DNA linear PAT 03-FEB-2004			
DEFINITION		Sequence 9293 from Patent WO02068579.					
ACCESSION		CQ723359					
VERSION		CQ723359.1 GI:42284216					
KEYWORDS		Homo sapiens (human)					
SOURCE		Homo sapiens					
ORGANISM		Homo sapiens					
		Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;					
		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
REFERENCE		1 Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.					
AUTHORS		Kits, such as nucleic acid arrays, comprising a majority of					
TITLE		humanexons or transcripts, for detecting expression and other uses					

QY	3016	CTCTTCTCTGAGTTCCCAAGACTCTAGCACTGGA	CTGTGGACCA	CAGCTCTCTGTGG	3075
Db	2397	CTCTTCTCTGAGTTCCCAAGACTCTAGCACTGGA	CTGTGGACCA	CAGCTCTCTGTGG	2456
QY	3076	GGGGAGGCTCTCTATCATCCCCAGTGTCTCAGG	CCGGGAAGCCG	GAAGTGA	CTGGCTAC 3135
Db	2457	GGGGAGGCTCTCTATCATCCCCAGTGTCTCAGG	CCGGGAAGCCG	GAAGTGA	CTGGCTAC 2516
QY	3136	TTCCCTCTGGGCATGTGACGACCTGACAGGCTG	CAGAGGTGCCAATAGAGCCCTTGGAGCCCTC	3195	
Db	2517	TTCCCTCTGGGCATGTGACGACCTGACAGGCTG	CAGAGGTGCCAATAGAGCCCTTGGAGCCCTC	2576	
QY	3196	CCACCCCACTGCTGACCTCCCGTGGAGCCGATC	CACAGCGAGGGG	CAGTGGGTGAGC	3255
Db	2577	CCACCCCACTGCTGACCTCCCGTGGAGCCGATC	CACAGCGAGGGG	CAGTGGGTGAGC	2636
QY	3256	CTGCGGGCCCCCTGACACCATCAAGTCCAGCTC	CGGGCTGGGTACATCAT	CCCCCTGC	3315
Db	2637	CTGCGGGCCCCCTGACACCATCAAGTCCAGCTC	CGGGCTGGGTACATCAT	CCCCCTGC	2696
QY	3316	CAGGGCTCTGGCTCAACACACAGAGTCCCGC	CAGCAGCCATGGCCCTGTGGCC	3375	
Db	2697	CAGGGCTCTGGCTCAACACACAGAGTCCCGC	CAGCAGCCATGGCCCTGTGGCC	2756	
QY	3376	CTAACCAAGGTGGAGAGGCCCGAGGGAGCTGT	TCTCGGACGATGGAGAGCCTGGAA	3435	
Db	2757	CTGACCAAGGTGGAGAGGCCCGAGGGAGCTGT	TCTCGGACGATGGAGAGCCTGGAA	2816	
QY	3436	GTGCTGGAGGAGGGCCCTACACAGGTCTATCT	TCTCGGACGATGGAGAGCCTGGAA	3495	
Db	2817	GTGCTGGAGGAGGGCCCTACACAGGTCTATCT	TCTCGGACGATGGAGAGCCTGGAA	2876	
QY	3496	AATGAGCTGTGAGTGTGACCAAGTGGAGCTG	CGCTCGAGCTGCAGAGGTGATGTC	3555	
Db	2877	AATGAGCTGTGAGTGTGACCAAGTGGAGCTG	CGCTCGAGCTGCAGAGGTGATGTC	2936	
QY	3556	CTGGCGTGGCCACCGCGCCCGACGAGTCTCT	CTCAACGGTGTCTCTCTCAACTTC	3615	
Db	2937	CTGGCGTGGCCACCGCGCCCGACGAGTCTCT	CTCAACGGTGTCTCTCTCAACTTC	2996	
QY	3616	ACCTAGAGCCCGACACCAAGTCTCTGACATCT	GTGTCTCGTGTGATGGAGAGCAG	3675	
Db	2997	ACCTAGAGCCCGACACCAAGTCTCTGACATCT	GTGTCTCGTGTGATGGAGAGCAG	3056	
QY	3676	TTTCTCTGCTGAGTGTGTT	3694		
Db	3057	TTTCTCTGCTGAGTGTGTT	3075		

RESULT 5	
HUMGAAA	
LOCUS	Human acid alpha-glucosidase (GAA) mRNA, complete cds.
DEFINITION	Human acid alpha-glucosidase (GAA) mRNA, complete cds.
ACCESSION	M34424
VERSION	M34424.1
KEYWORDS	acid alpha-glucosidase; acid maltase.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	Martiniuk, F., Mehler, M., Tzall, S., Meredith, G. and Hirschhorn, R.
TITLE	Sequence of the cDNA and 5'-flanking region for human acid alpha-glucosidase, detection of an intron in the 5' untranslated leader sequence, definition of 18-bp polymorphisms, and differences with previous cDNA and amino acid sequences
JOURNAL	DNA 9, 85-94 (1990)
COMMENT	Original source text: Human, cDNA to mRNA.
FEATURES	Location/Qualifiers
source	1..3441
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"

gene	/map="17q23"
CDS	1..3441
	/gene="GAA"
	33..2891
	/gene="GAA"
	/note="acid alpha-glucosidase"
	/codon_start=1
	/protein_id="AAA52506.1"
	/db_xref="GI:182908"
	/db_xref="GDB:G00-119-965"
	/translation="MGVHRPCSHRLAVCALVSLATAALLGHILLHDFLLVPRLSGSSPVEETHPAHQGASRPRDQAHPRAVPTQCDVPNSRFDCAPDKAITQEQCEAGCCYIPAKQGLQGMQPCFFPSPYKLENLSSEMSQITLRTTPTTFPKDILTLRDMVTENRHTIKDPAANRVEVPLETHPHSRAPSPLYSVFSEBEPFVIVRDLQDLNNTVAFLFFADQFLQSLTSPQVITGLAEHLSPLMLSTWTRITLNRDLAPTGANLYGSHPFYLALEDGSGAHGVLNANMDVLPQSPALSWSTGGLDYVIFLGPBPKSVQVQYLDVGVYPMPYPWGLGHLCRNGYSSTAITRQVVENMTRAHFDVQWDLNDYMSRRDFTFNKGRDFFPAMVQELHQGGRRYMTVDPAISSSGPAGSYRYPDEGLRRGVFTINETGQPLIGKWPSTAFPDPFTNPTALAWMDVAFHDSQVFPDGMWIDMNEPSNFIIRGSDGCPNNELENPPYVGVGTGTTQAATICASHQFSLTHYNLLYGLTEAISHRALVKARTRPFVLSRSTFAGHRYAGHWTGDMVMSWEQLASSVPEILQFNLLGVPLVADVCVCGPLGNTSELCVWTOLGAFVPMENHNSLLSLPQEPYSEFPAQAMRKALNRLRYALLPHLYTLFQHAHVAGETVARPLFLRPPKDSSTWTVHQLLMGALLITPVLQAGKAETVGYFFPLGTWYDLQVTFIEALGSLATPPAPRPAIHSQGWTLPLAPLDTINVHLRAGYIPLQGFLLTTSRQPPMALVALTKGGEARGE LFWDDGESLEVLERGAYTVIFLARNTIVNELVRVTSEGAGLQQLQKVTVLGATAPQ QVLSNGVPVSNFTYSPDTKVLIDICVSLLMGQFLVSWC"

ORIGIN

Query Match	71.3%	Score	2639.8	DB	9	Length	3441
Best Local Similarity	99.5%	Pred. No.	0				
Matches	2647	Conservative	0	Mismatches	12	Indels	0
Gaps	0						
QY	1036	GAACCGGTGCACACCCCGCGCTCCAGAGCAGTCCACACAGTGGAGCTGCCCC	1095				
Db	231	GATGCCAGGCACACCCCGCGCTCCAGAGCAGTGGAGCTGCCCC	290				
QY	1096	AACAGCCGCTTCGATTGCGCCCTCGAAGGCCATCACCCAGGAACAGTGCAGGCGCCG	1155				
Db	291	AACAGCCGCTTCGATTGCGCCCTCGAAGGCCATCACCCAGGAACAGTGCAGGCGCCG	350				
QY	1156	GGCTGCTGATCATCCCTGCAAGCAGGGGTGAGGAGCCAGATGGGCGAGCCCTGG	1215				
Db	351	GGCTGCTGATCATCCCTGCAAGCAGGGGTGAGGAGCCAGATGGGCGAGCCCTGG	410				
QY	1216	TGCTTCTTCCACCCAGCTACCCAGCTACAAGCTGAGAGAACCTGAGCTCTCTGAATG	1275				
Db	411	TGCTTCTTCCACCCAGCTACCCAGCTACAAGCTGAGAGAACCTGAGCTCTCTGAATG	470				
QY	1276	GGCTACAGCGCCACCTGACCCGCTACCAACCCGCTTCTTCTTCCCAAGGACATCTGACC	1335				
Db	471	GGCTACAGCGCCACCTGACCCGCTACCAACCCGCTTCTTCTTCCCAAGGACATCTGACC	530				
QY	1336	CTGCGGCTGAGCGTATGATGAGAGCTGAGAACCCGCTCCACTTCAGATCAAGATCCA	1395				
Db	531	CTGCGGCTGAGCGTATGATGAGAGCTGAGAACCCGCTCCACTTCAGATCAAGATCCA	590				
QY	1396	GCTACAGCGCGTACAGGTGCTTGGAGACCCCGCTGTCACAGCGGCGCACCGTCC	1455				
Db	591	GCTACAGCGCGTACAGGTGCTTGGAGACCCCGCTGTCACAGCGGCGCACCGTCC	650				
QY	1456	CCACTCTACAGCGTGGAGTCTTCCGAGGAGCCCTTCCGGGGTGATCGTCACCGGAGCTG	1515				
Db	651	CCACTCTACAGCGTGGAGTCTTCCGAGGAGCCCTTCCGGGGTGATCGTCACCGGAGCTG	710				
QY	1516	GACGCGCGGCTGCTGCTGAAACAGCGGCTGGCGCCCTGTTCTTTTGGGACAGTTCCTT	1575				
Db	711	GACGCGCGGCTGCTGCTGAAACAGCGGCTGGCGCCCTGTTCTTTTGGGACAGTTCCTT	770				
QY	1576	CAGCTGTCCACTCGCTCGCTCGAGTATATACAGGCGCTCGCGGAGCACCTCAGTCC	1635				
Db	771	CAGCTGTCCACTCGCTCGCTCGAGTATATACAGGCGCTCGCGGAGCACCTCAGTCC	830				

Qy	1636	CTGATGCTCAGCACACAGCTGGACAGAGATCACCTGTGGAAACCGGAGACTTTCGGCCCAAG	1699
Db	831	CTGATGCTCAGCACACAGCTGGACAGAGATCACCTGTGGAAACCGGAGACTTTCGGCCCAAG	890
Qy	1696	CCGGTTCGGAACCTCTACGGGTCTCACCTTTTCTACCTGGCGCTGGAGGACGGCGGGTTCG	1755
Db	891	CCGGTTCGGAACCTCTACGGGTCTCACCTTTTCTACCTGGCGCTGGAGGACGGCGGGTTCG	950
Qy	1756	GCACACGGGGTGTTCCTGCTTAAACAGACAAATGCACATGGATGTGTCTCTGACGCGAGCCCT	1815
Db	951	GCACACGGGGTGTTCCTGCTTAAACAGACAAATGCACATGGATGTGTCTCTGACGCGAGCCCT	1010
Qy	1816	GCCTTAGCTGGAGGTTCGACAGGTGGGATCCTTGGATGTCTACATCTTCCTGGGCCAGAG	1875
Db	1011	GCCTTAGCTGGAGGTTCGACAGGTGGGATCCTTGGATGTCTACATCTTCCTGGGCCAGAG	1070
Qy	1876	CCCNAGAGCGTGGTGCAGCAGTACCTTGGACGTTGTGGGATACCGGTTCATGCGGCCATAC	1935
Db	1071	CCCNAGAGCGTGGTGCAGCAGTACCTTGGACGTTGTGGGATACCGGTTCATGCGGCCATAC	1130
Qy	1936	TGGGGCTTGGCTTTCACCTGTGCGCGTGGGGTACTTCCTCCACCGCTATACCCGCCAG	1995
Db	1131	TGGGGCTTGGCTTTCACCTGTGCGCGTGGGGTACTTCCTCCACCGCTATACCCGCCAG	1190
Qy	1996	GTGTGGAGAACATGACACAGGGGCCACTTCCCCTTGGACGTCCAAATGGAAAGACCTGGAC	2055
Db	1191	GTGTGGAGAACATGACACAGGGGCCACTTCCCCTTGGACGTCCAAATGGAAAGACCTGGAC	1250
Qy	2056	TACATGSACTCCCGGAGGACTTCACGTTCAAACAGGATGGCTTCCGGAGCTTCCCGGCC	2115
Db	1251	TACATGSACTCCCGGAGGACTTCACGTTCAAACAGGATGGCTTCCGGAGCTTCCCGGCC	1310
Qy	2116	ATGTGTGACGAGCTGCACACAGGGCGGCGCGGCTACATGATGATCGTGGATCCTGGCATC	2175
Db	1311	ATGTGTGACGAGCTGCACACAGGGCGGCGCGGCTACATGATGATCGTGGATCCTGGCATC	1370
Qy	2176	AGCAGCTCGGGCCCTGCGCGGAGCTACAGGCCCTTACGACGAGGGTCTGCGGAGGGGGTT	2235
Db	1371	AGCAGCTCGGGCCCTGCGCGGAGCTACAGGCCCTTACGACGAGGGTCTGCGGAGGGGGTT	1430
Qy	2236	TTTATACACACGAGACCGGCGCAGCCGCTGATTTGGGAGGTATGGCCCGGGTCCACTGCC	2295
Db	1431	TTTATACACACGAGACCGGCGCAGCCGCTGATTTGGGAGGTATGGCCCGGGTCCACTGCC	1490
Qy	2296	TTCCCGGACTTACCAACCCACACAGCCCTTGGCTGTGGAGGACATGTGTGGCTGAGTTTC	2355
Db	1491	TTCCCGGACTTACCAACCCACACAGCCCTTGGCTGTGGAGGACATGTGTGGCTGAGTTTC	1550
Qy	2356	CATGACAGGTGCCCTTTCGACGGCTGTGATTGACATGAACGAGCCTTCCAACTTCAATC	2415
Db	1551	CATGACAGGTGCCCTTTCGACGGCTGTGATTGACATGAACGAGCCTTCCAACTTCAATC	1610
Qy	2416	AGAGGCTCTGAGGACGGCTGCCCCAAACAATGAGCTGGAGAAACCAACCTACGTGCTGGG	2475
Db	1611	AGAGGCTCTGAGGACGGCTGCCCCAAACAATGAGCTGGAGAAACCAACCTACGTGCTGGG	1670
Qy	2476	GTGTTTGGGGGACCCCTTCACAGGGGCCACCATCTGTGCTCTCCAGCCACCAAGTTTCTCTCC	2535
Db	1671	GTGTTTGGGGGACCCCTTCACAGGGGCCACCATCTGTGCTCTCCAGCCACCAAGTTTCTCTCC	1730
Qy	2536	ACACATCAACCTTGCAAACTCTACGGCTTCACGGCTTCGAAAGCCATCGCCTCCCAAGGGG	2595
Db	1731	ACACATCAACCTTGCAAACTCTACGGCTTCACGGCTTCGAAAGCCATCGCCTCCCAAGGGG	1790
Qy	2596	CTGGTGAAGCTCGGGGACACGGCCATTTGTGATCTCCCGCTCGACCTTGGCTGGGCCAC	2655
Db	1791	CTGGTGAAGCTCGGGGACACGGCCATTTGTGATCTCCCGCTCGACCTTGGCTGGGCCAC	1850
Qy	2656	GGCCGATACCGCGGCACCTGGAAGGGGACGTGTGGAGCTCCTGGGAGCAGCTCGCTTCC	2715
Db	1851	GGCCGATACCGCGGCACCTGGAAGGGGACGTGTGGAGCTCCTGGGAGCAGCTCGCTTCC	1910
Qy	2716	TCCGTGCCAGAAATCTCTGCATTTAACTCCTCGGGGTGCTCTGTGTGGGGCCGACGTC	2775

RESULT 6
BC040431
LOCUS
DEFINITION

IMAGE:5739991), complete cds.	
BC040431	/codon_start=1
VERSION	/product="GAA protein"
KEYWORDS	/protein_id="AAH0431.1"
SOURCE	/db_xref="GI:26251857"
ORGANISM	/db_xref="MIM:606800"
	/db_xref="LocusID:2548"
	/db_xref="MIM:606800"
REFERENCE	/translation="MGVRHPPCSHRLAVCALVSLAALLGHIILLHDFLLVPRELSG
AUTHORS	SSPVLTEHPAQOQASRPGPRDAQHPRAVPTQCDVPPNSRFDCAIDKAIQEQ
	CEARCCVYIPAKOGLOAGMGOPWCFFPPSPYSKLENLSSEMGGYATLTTRTTFP
	PKDILTLDVMMETENLHPTIKDPANRYEVPLETPHVHSRPSPLYSVFSREPF
	GVIVRQDGRVLLNTVAPLFFADQFLQSTLSFSQITGLAHLSPMLSTSTRI
	TLMKNDLPTGANLYGHPFLALEDGSGAHGFLNSNAMVVLPQSPPLSLSPQ
	GILDYIFLGPSPSVQYLDVGVGPPMPYGLGFLHCRMGYSTAIRQVSNMT
	RAHFLPYDQNDLDRDFTFNKQDFRPPAMVQBLHOGGRYMMIVDAISSG
	PAGSVYDEGLRGVFTTNGTGLIGKWPFGSTAFDFTNPTALAWEDMVAEFLS
	QVPPDMYDWNENSFIRGSDGCPNNELENPPYVGVGCTLOAATCASHOFLS
	THYNLLHGLTEALASHALVKARTRPFRVLSRSTFAGHRYAGHNTGDWNSWEQL
	ASSVPEILQFNLLGVPLVAGDGFGLNTSELVCRVWTOGAFYFPMNRHNSLLSPQ
	EPYSFEPAQQAQMRKALTRYALLPHLYTLFQAHVAGETVARPLFLFPKDSSTWTY
	DHQLIWMGALLITPVLOAKAEVGYFPLGTWYDQTVFVEALGPLPPPPAPRPAI
	HSEGWVTLPAPLDTINVHLRAGYIIPLOGPLTITTESRQOQPMALAVLTKEGEARGE
	LFWDGSELEVLERGAYTVLFLARNNTIVNELVAVTSEGAGLQKQKTVLGVATAPO
	QVLSNGVPVSNFTVSPDKVLIDICVSLLMGQFLVSWC"
	385..546
misc_feature	/gene="GAA"
	/note="PD; Region: P or trefoil or TFF domain"
	/db_xref="CDD:smart00018"
misc_feature	724..2895
	/gene="GAA"
	/note="Glyco_hydro_31; Region: Glycosyl hydrolases family
	31. Glycosyl hydrolases are key enzymes of carbohydrate
	metabolism. Family 31 comprises of enzymes that are, or
	similar to, alpha-galactosidases"
	/db_xref="CDD:pfam01055"
ORIGIN	
	Query Match 71.2%; Score 2635; DB 9; Length 3687;
	Best Local Similarity 99.4%; Pred. No. 0;
	Matches 2644; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY	1036 GAAACGGTGCACACCCCGCGTCCAGAGCAGTGCACACAGTGCAGCTCCCCC 1095
DB	
	343 GATGCCAGGCACACCCCGCGTCCAGAGCAGTGCACACAGTGCAGCTCCCCC 402
QY	1096 AACACCGCTTCGATTGCGCCCTGCACAGGCGCATCCACAGGAACAGTGCAGGCGCCG 1155
DB	
	403 AACACCGCTTCGATTGCGCCCTGCACAGGCGCATCCACAGGAACAGTGCAGGCGCCG 462
QY	1156 GGCTCTGTACATCCCTGCAAGCAGGGGTGTCAGGAGGCCAGATGGGGCAGCCCTGG 1215
DB	
	463 GGCTGTGTACATCCCTGCAAGCAGGGGTGTCAGGAGGCCAGATGGGGCAGCCCTGG 522
QY	1216 TGCTTCTTCCACCCAGCTACCCAGCTACAGCTGGAGAACCTGAGCTCTCTGAAATG 1275
DB	
	523 TGCTTCTTCCACCCAGCTACCCAGCTACAGCTGGAGAACCTGAGCTCTCTGAAATG 582
QY	1276 GGCTACAGCGGCACCCCTGACCCGTACACCCGCCCTTCTTCCCAAGGACATCCTGACC 1335
DB	
	583 GGCTACAGCGGCACCCCTGACCCGTACACCCGCCCTTCTTCCCAAGGACATCCTGACC 642
QY	1336 CTGGCGCTGGAGCTGATGATGGAGACTGAGAACCCGCCCTCCACTTCACGATCAAGATCCA 1395
DB	
	643 CTGGCGCTGGAGCTGATGATGGAGACTGAGAACCCGCCCTCCACTTCACGATCAAGATCCA 702
QY	1396 GCTACAGCGCTACAGCTGCGCTTGGAGACCCCGCTGTCACAGCGCGGCACCGCTCC 1455
DB	
	703 GCTAACAGCGCTACAGCTGCGCTTGGAGACCCCGCTGTCACAGCGCGGCACCGCTCC 762
QY	1456 CCACCTCTACAGCGTGGAGTTCTCCGAGGAGCCCTTTCGGGGTGATCGTGCACCGGAGCTG 1515
DB	
	763 CCACCTCTACAGCGTGGAGTTCTCCGAGGAGCCCTTTCGGGGTGATCGTGCACCGGAGCTG 822
QY	1516 GACGGCGCGTGTCTGCTGAACACAGAGGTGGCGCCCTTGTCTTTTTCGGGAACAGTTCCTT 1575

Db	823	GACGGCCGCGTGTCTGTGAAACAAGCAGGTGGCGCCCTGTTCCTTTCGGGACCAAGTTCCTTT	882
Qy	1576	CAGCTGTCCACCTCGCTCGCCCTCGCAGTATATACACAGGCGCTCGCCGAGCACCTCAGTCCC	1635
Db	883	CAGCTGTCCACCTCGCTCGCCCTCGCAGTATATACAGGCGCTCGCCGAGCACCTCAGTCCC	942
Qy	1636	CTGATGCTCAGCACACAGCTGGACAGAGATCACCTGTGGAACCGGACCTTTCGCCCAAG	1695
Db	943	CTGATGCTCAGCACACAGCTGGACAGAGATCACCTGTGGAACCGGACCTTTCGCCCAAG	1002
Qy	1696	CCGGTTCGGAACCTCTACGGGTCTCACCCCTTTCTACCTGGCGCTGGAGAGCGGGGTGG	1755
Db	1003	CCGGTTCGGAACCTCTACGGGTCTCACCCCTTTCTACCTGGCGCTGGAGAGCGGGGTGG	1062
Qy	1756	GCAACACGGGTGTTCTGCTTAAACAGCAATGCCATGGATGTGGTCTCTGCAGCCGAGCCCT	1815
Db	1063	GCAACACGGGTGTTCTGCTTAAACAGCAATGCCATGGATGTGGTCTCTGCAGCCGAGCCCT	1122
Qy	1816	GCCTTAGCTGGAGGTGCACAGGTGGGATCCTGGATGTCTACATCTTCTTGGGCCAGAG	1875
Db	1123	GCCTTAGCTGGAGGTGCACAGGTGGGATCCTGGATGTCTACATCTTCTTGGGCCAGAG	1182
Qy	1876	CCCAAGAGCGTGTGCAGCATGCTTGGAGCTTGTGGGATACCGGTTCATGCGCGCATAC	1935
Db	1183	CCCAAGAGCGTGTGCAGCATGCTTGGGATACCGGTTCATGCGCGCATAC	1242
Qy	1936	TGGGGCTGGGCTTCACACCTGTGCGGTGGGGCTACTCTCCACCGCTATACCCGCCAG	1995
Db	1243	TGGGGCTGGGCTTCACACCTGTGCGGTGGGGCTACTCTCCACCGCTATACCCGCCAG	1302
Qy	1996	GTGTTGAGAAACATGACAGGGGCCACTTCCCTCGACGTCCAATGGAAAGCACTGGAC	2055
Db	1303	GTGTTGAGAAACATGACAGGGGCCACTTCCCTCGACGTCCAATGGAAAGCACTGGAC	1362
Qy	2056	TACATGACTCCCGGAGGGACTTACGTTCAAAGGATGGCTTCGGGACTTCCCGGCC	2115
Db	1363	TACATGACTCCCGGAGGGACTTACGTTCAAAGGATGGCTTCGGGACTTCCCGGCC	1422
Qy	2116	ATGTTGACAGCTGCACAGGGCGCGCGGTACATGATGATCGTGGATCCTGCCATC	2175
Db	1423	ATGTTGACAGCTGCACAGGGCGCGCGGTACATGATGATCGTGGATCCTGCCATC	1482
Qy	2176	AGCAGCTTCGGGCCCTGCGGGGAGCTACAGGCCCTACGACAGGGTCTGCGAGGGGGGT	2235
Db	1483	AGCAGCTTCGGGCCCTGCGGGGAGCTACAGGCCCTACGACAGGGTCTGCGAGGGGGGT	1542
Qy	2236	TTCATACCAAAGAGACCGGCGAGCGCTGATTGGGAAGGTATGGCCCGGGTCCACTGCC	2295
Db	1543	TTCATACCAAAGAGACCGGCGAGCGCTGATTGGGAAGGTATGGCCCGGGTCCACTGCC	1602
Qy	2296	TTCCCGACTTCAACCAACCCACAGCCCTGGCTGGGAGGACATGGTGGCTGAGTTC	2355
Db	1603	TTCCCGACTTCAACCAACCCACAGCCCTGGCTGGGAGGACATGGTGGCTGAGTTC	1662
Qy	2356	CATGACAGGTGCCCCCTTCGACGGCTTGTGATTGACATGAACAGAGCTTCCAACTTCATC	2415
Db	1663	CATGACAGGTGCCCCCTTCGACGGCNATGGATGACATGAACAGAGCTTCCAACTTCATC	1722
Qy	2416	AGAGGCTCTGAGGACGGCTGCCCAACAATGAGCTGGAGAACCCAACCTTACGTGCTGGG	2475
Db	1723	AGGSGCTCTGAGGACGGCTGCCCAACAATGAGCTGGAGAACCCAACCTTACGTGCTGGG	1782
Qy	2476	GTGGTGGGGGACCTTCAGGGGGCCACCATCTGTGCTCTCAGCCACCAAGTTCTCTCC	2535
Db	1783	GTGGTGGGGGACCTTCAGGGGGCCACCATCTGTGCTCTCAGCCACCAAGTTCTCTCC	1842
Qy	2536	ACACACTACAACTGCACAACTCTACGGCTTGACCGAAGCCATCGCTCCACAGAGGGG	2595
Db	1843	ACACACTACAACTGCACAACTCTACGGCTTGACCGAAGCCATCGCTCCACAGAGGGG	1902
Qy	2596	CTGGTGAAGGCTCGGGGGAACGCCCCATTTGTGATCTCCGCTCGACCTTTGTGTGCCAC	2655

Db	1903	CTGGTGAAGGCTCGGGGGACACGCCCATTTGTGATCTCCGCTCGACACTTTGCTGCGCCAC	1966
Qy	2656	GGCCGATACGCCCGGCCCATCTG3ACGGGG3ACGTGTGGAGCTCTCTGG3AGACAGCTCGCCTCC	2715
Db	1963	GGCCGATACGCCCGGCCCATCTG3ACGGGG3ACGTGTGGAGCTCTCTGG3AGACAGCTCGCCTCC	2022
Qy	2716	TCGGTGC3AGAAATCTCGAGATTTAACTGTCTGGGGGTGCCTCTGTCTCGGGGCCGAGCTC	2775
Db	2023	TCGGTGC3AGAAATCTCGAGATTTAACTGTCTGGGGGTGCCTCTGTCTCGGGGCCGAGCTC	2082
Qy	2776	TGCGGCTTCTCGGGCAACACCTCAGAGGAGCTGTGTGCGCTGG3ACCCAGCTTGGGGGCC	2835
Db	2083	TGCGGCTTCTCGGGCAACACCTCAGAGGAGCTGTGTGCGCTGG3ACCCAGCTTGGGGGCC	2142
Qy	2836	TTCTACCCCTTCATGCG3AACACAAACAGCCTGCTCAGTCTGCCCCAGGAGCGGTACAGC	2895
Db	2143	TTCTACCCCTTCATGCG3AACACAAACAGCCTGCTCAGTCTGCCCCAGGAGCGGTACAGC	2202
Qy	2896	TTCAGGAGCGCGCCACAGAGGCCATGAG3AGGCCCTCACCTTGGGCTTACGCACTCCTC	2955
Db	2203	TTCAGGAGCGCGCCACAGAGGCCATGAG3AGGCCCTCACCTTGGGCTTACGCACTCCTC	2262
Qy	2956	CCCCACTCTACACACTGTTTCACACAGGCCACAGTCGCGGGGG3AGACCGTGGGCCCGGCC	3015
Db	2263	CCCCACTCTACACACTGTTTCACACAGGCCACAGTCGCGGGGG3AGACCGTGGGCCCGGCC	2322
Qy	3016	CTCTCTCGAGTTCCCCAAAGACTCTAGCACCTG3ACTGTGTG3ACCAACAGCTCTCTGTGG	3075
Db	2323	CTCTCTCGAGTTCCCCAAAGACTCTAGCACCTG3ACTGTGTG3ACCAACAGCTCTCTGTGG	2382
Qy	3076	GGGGAGGCCCTGCTCATACACCCAGTCTC3AGGCCGG3AGAGGCCCGAAGT3AGTGGCTAC	3135
Db	2383	GGGGAGGCCCTGCTCATACACCCAGTCTC3AGGCCGG3AGAGGCCCGAAGT3AGTGGCTAC	2442
Qy	3136	TTCCCTTGGGCACATG3TACGACCTG3CAGACGGTCCAAATAG3AGGCCCTTGG3AGCCTC	3195
Db	2443	TTCCCTTGGGCACATG3TACGACCTG3CAGACGGTCCAAATAG3AGGCCCTTGG3AGCCTC	2502
Qy	3196	CCACCCCACTTGCAGCTCCCGT3AGCCAGCCAGCCATCCACAGCGAGGGG3AGTGGGTGACG	3255
Db	2503	CCACCCCACTTGCAGCTCCCGT3AGCCAGCCAGCCATCCACAGCGAGGGG3AGTGGGTGACG	2562
Qy	3256	CTGCGCGGCCCTTG3ACACCAATCAAAGTTC3ACCTCCGGGCTGG3TACATCATCCCCCTG	3315
Db	2563	CTGCGCGGCCCTTG3ACACCAATCAAAGTTC3ACCTCCGGGCTGG3TACATCATCCCCCTG	2622
Qy	3316	CAGGGCCCTGGCTCACAAC3ACAGAGTCCCG3ACAGCCCATGGCCCTGGCTGTGGCC	3375
Db	2623	CAGGGCCCTGGCTCACAAC3ACAGAGTCCCG3ACAGCCCATGGCCCTGGCTGTGGCC	2682
Qy	3376	CTAAACAAAGGTTGGAGAGCCCGAGGG3AGCTGTTCTGG3ACGATGGAGAGAGCCTGGAA	3435
Db	2683	CTGACCAAGGTTGGAGAGCCCGAGGG3AGCTTTCCTGG3ACGATGGAGAGAGCCTGGAA	2742
Qy	3436	GTGCTGGAGCGAGGGCCCTACACAG3TCACTTCTTG3CCAG3AATTAACAGATCGTG	3495
Db	2743	GTGCTGGAGCGAGGGCCCTACACAG3TCACTTCTTG3CCAG3AATTAACAGATCGTG	2802
Qy	3496	AATGAGCTGGTAGTGT3ACCAAGTGG3AGCTGGCCCTG3AGCCATGGCCCTGGCTGTGGCC	3555
Db	2803	AATGAGCTGGTAGTGT3ACCAAGTGG3AGCTGGCCCTG3AGCCATGGCCCTGGCTGTGGCC	2862
Qy	3556	CTGGGCTGGCCACGGCGCC3CCAGCAG3TCCCTCTCCAA3GGTGTCCCTGTCTCCAACTTC	3615
Db	2863	CTGGGCTGGCCACGGCGCC3CCAGCAG3TCCCTCTCCAA3GGTGTCCCTGTCTCCAACTTC	2922
Qy	3616	ACCTACAGCCCC3ACCAAG3TCTTGG3ACATCTGTGTCTCGCTGTGTATGG3AGAGCAG	3675
Db	2923	ACCTACAGCCCC3ACCAAG3TCTTGG3ACATCTGTGTCTCGCTGTGTATGG3AGAGCAG	2982
Qy	3676	TTTCTCGTACGTGGTGT	3694
Db	2983	TTTCTCGTACGTGGTGT	3001

Db 1710 CACACACTAACCTGACAAACCTCTACGGCCCTGACCGAAGCATCGCTCCACAGGGC 1651
Qy 2595 GCTGGTGAAGGCTCGGGGACACGCCCATTTGTGATCTCCCGCTCGACCTTTCTCGGCCA 2654
Db 1650 GCTGGTGAAGGCTCGGGGACACGCCCATTTGTGATCTCCCGCTCGACCTTTCTCGGCCA 1591
Qy 2655 CGCCCGATACCGCGGCACATGAGACGGGGAGCGTGTGAGAGCTCTCGGAGAGAGCTCGCCTC 2714
Db 1590 CGCCCGATACCGCGGCACATGAGACGGGGAGCGTGTGAGAGCTCTCGGAGAGAGCTCGCCTC 1531
Qy 2715 CTCCTGCGCAAAATCTCGAGTTTAACTGCTGGGGGTCCTCTGTCGGGGCCGACGT 2774
Db 1530 CTCCTGCGCAAAATCTCGAGTTTAACTGCTGGGGGTCCTCTGTCGGGGCCGACGT 1471
Qy 2775 CTCGGGCTTCTCGGGCAACACTCAGAGGAGCTGTGTGTCGCTGGACCCAGCTGGGGGC 2834
Db 1470 CTCGGGCTTCTCGGGCAACACTCAGAGGAGCTGTGTGTCGCTGGACCCAGCTGGGGGC 1411
Qy 2835 CTTCTACCCCTTATCGGGGAACACAAAGCCTGCTCAGTGTGCCCGCAGAGCCGTACAG 2894
Db 1410 CTTCTACCCCTTATCGGGGAACACAAAGCCTGCTCAGTGTGCCCGCAGAGCCGTACAG 1351
Qy 2895 CTTCAGGAGCGCGCCAGCAGGCCATGAGAGGCCCTCACCTGGCTACGCACTCCT 2954
Db 1350 CTTCAGGAGCGCGCCAGCAGGCCATGAGAGGCCCTCACCTGGCTACGCACTCCT 1291
Qy 2955 CCCCCACTCTACACACTGTTCCACAGGCCACGTCGCGGGGAGACCGTGGCCCGGCC 3014
Db 1290 CCCCCACTCTACACACTGTTCCACAGGCCACGTCGCGGGGAGACCGTGGCCCGGCC 1231
Qy 3015 CTTCTTCCTGAGTTCCCAAGAGCTCTAGCAGCTGGAAGAGCCCTCACCTGGCTACGCACTCCTGTG 3074
Db 1230 CTTCTTCCTGAGTTCCCAAGAGCTCTAGCAGCTGGAAGAGCCCTCACCTGGCTACGCACTCCTGTG 1171
Qy 3075 GGGGAGGCGCTCTCATACCCAGCTGCTCAGGCGCGGAGAGCCGAGTGAAGTGAAGTGA 3134
Db 1170 GGGGAGGCGCTCTCATACCCAGCTGCTCAGGCGCGGAGAGCCGAGTGAAGTGAAGTGA 1111
Qy 3135 CTTCCCTTTGGGCACATGCTAGCAGCTGCGAGCGTGCCTCAATAGAGCCCTTGGCAGCCT 3194
Db 1110 CTTCCCTTTGGGCACATGCTAGCAGCTGCGAGCGTGCCTCAATAGAGCCCTTGGCAGCCT 1051
Qy 3195 CCACCCCACTGAGTCTCCCGTGAAGCAGCAGCTCCACAGCGAGGGGAGTGGGTGAC 3254
Db 1050 CCACCCCACTGAGTCTCCCGTGAAGCAGCAGCTCCACAGCGAGGGGAGTGGGTGAC 991
Qy 3255 GCTGCGGCGCCCTCGACACATCAACCTGCACTCCGCGCTGCGGTGCTATCATATCCCT 3314
Db 990 GCTGCGGCGCCCTCGACACATCAACCTGCACTCCGCGCTGCGGTGCTATCATATCCCT 931
Qy 3315 GCAGGCGCTTGGCTCAACACACAGAGTCCCGCCAGAGCCCATGGCCCTGGCTGTGGC 3374
Db 930 GCAGGCGCTTGGCTCAACACACAGAGTCCCGCCAGAGCCCATGGCCCTGGCTGTGGC 871
Qy 3375 CTTAACAAAGGTGAGAGCCCGAGGGAGAGTGTCTGGAGCAGTGGAGAGAGCTTGA 3434
Db 870 CTTGACCAAGGTGAGAGCCCGAGGGAGAGTGTCTGGAGCAGTGGAGAGAGCTTGA 811
Qy 3435 AGTGTGAGAGCGAGGGCCCTACACAGGTTCATCTTCCTGGCCAGAGATTAACAGATCT 3494
Db 810 AGTGTGAGAGCGAGGGCCCTACACAGGTTCATCTTCCTGGCCAGAGATTAACAGATCT 751
Qy 3495 GAATGAGCTGGTACGTGTACACAGTGAAGGAGCTGGCCCTGCAGCTGCAG-AAAGTGAAGT 3553
Db 750 GAATGAGCTGGTACGTGTACACAGTGAAGGAGCTGGCCCTGCAGCTGCAGCAAGTGAAGT 691
Qy 3554 TCTTGGGCGTGGCCAAGGCGCCCAAGAGTCTCTCTCAAGCGTGTCTCTTCTCAACT 3613
Db 690 TCTTGGGCGTGG--CAGGCGCCCAAGAGTCTCTCTCAAGCGTGTCTCTTCTCAACT 632
Qy 3614 TCACCTACAGCCCGGACACCAAGTCTTGGACATCTGTCTGCTGTGTGATGGGAGAGC 3673
Db 631 TCACCTACAGCCCGGACACCAAGTCTTGGACATCTGTCTGCTGTGTGATGGGAGAGC 572

Qy 3674 AGTTTCTCTCAGCTGGTGT 3694
Db 571 AGTTTCTCTCAGCTGGTGT 551

RESULT 8
AF171665 2814 bp mRNA linear MAM 04-JUL-2000
LOCUS Bos taurus acidic alpha-glucosidase mRNA, complete cds.
DEFINITION
AF171665
ACCESSION
AF171665.1 GI:8925837
VERSION
KEYWORDS
SOURCE
Bos taurus (cow)
ORGANISM
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Ruminantia; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bos.
REFERENCE
1 (bases 1 to 2814)
Dennis,J.A., Moran,C. and Healy,P.J.
The bovine alpha-glucosidase gene: coding region, genomic
structure, and mutations that cause bovine generalized glycosenosis
Mamm. Genome 11 (3), 206-212 (2000)
20188772
MEDLINE
10723725
REFERENCE
2 (bases 1 to 2814)
Dennis,J.A., Moran,C. and Healy,P.J.
Direct Submission
Submitted (22-JUL-1999) Agriculture, Elizabeth Macarthur
Agricultural Institute, Woodbridge Road, Menangle, New South Wales
2568, Australia
Location/Qualifiers
1..2814
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/cell_type="fibroblast"
1..2814
/note="AAG"
/codon_start=1
/product="acidic alpha-glucosidase"
/protein_id="AAF81636.1"
/db_xref="GI:8925838"
/translation="MMRWPPCPSPRLPLGVLTLLSLALLHLHLHLEWVPREURGFSD
EIHQACQPCASSPEGRCPRAAPQCDLPNSRFDPCAPDKGITPOCEARGCCMPAE
WPDQMGOPWCFFPSPYSYLENLTETGTATLTTRAVTFFPKDITMLRLDMLM
ETESLHPTIKDPAERREVEPLETRVYSQAPPLYSVESFSEPGVVRRLDGRVL
LNTVAPFPADQFQLSTLSQSHITGLAHLGSLMLSTNWKITLNRDITAPENV
NLYGHPFYLVEGDGLAHGVFLNSNMDVVLQSPALUSWRSTGGILDVYIFLGPPEP
KSVVQYLDVVGYPFMPYWGILGPHLCRMGYSTAITROVENMTRAYFLDVQWDL
DYMDARRDPTFNKHGDFPAMVQELHOGGRYIMVDPAISSGPGATYRYPDEGLR
RGVFTNTEGQPLIGQVWPGLTAFDPFTNPETLDWQDMVTFHAQVDFDGMWIDNE
PSNFVSGSDGCPDNLSPYLPVGVGTLRAATICASSHQFLSTHYDLHLNGLTE
ALASHRALVKARCMRPFVLSRSTFAGHRYSGHWGTDVWSNWEQLSYSPVETLLFNLL
GVPLGADICGLGNTSBEICVRVGTQALGFYFPMRNNALNSQPPYRPFSETAQAM
RKAFTRVLLPYLTLFRAHVRGETVARPLFEPPEDPTWTVDRLMLGEALLIT
PVLEAKVEVTGFPQGTWYDQTVMEAFGLSPPLPLSVTHSGQWVTSAPLDT
INVHLAGHII PMQGPALTTSRKOHMALAVALTASGEAQLFWDGSESLQVLDGG
DYTQLIFLAKNTFUNKLVHVSSEGASLQLRNVTVLGVATAPQQLCNSVPVSNFTFS
PDTETLAIPLVSLTMGEQFVSHS"

Query Match 54.9%; Score 2033.2; DB 4; Length 2814;
Best Local Similarity 86.0%; Pred. No. 2,6e-305;
Matches 2280; Conservative 0; Mismatches 363; Indels 9; Gaps 2;
Qy 1045 GCACACCCCGCGCTCCACAGAGAGTGGCCACACAGTGGCAGCTCCCCAACACGCCG 1104
Db 172 GAATGCGGTGGCAGGCCCGCCAGGCGAGCCTACACAGTGGCAGCTCGCCCCAACACGCCG 231
Qy 1105 TTTCGATTGGCCCTGACAAAGCCATCACCCAGGACAGTGGAGCCCGCGCTCTGCTGC 1164
Db 232 TTTCGATTGGCCCTGACAAAGCCATCACCCAGGAGAGTGGAGCCCGCTCTGCTGC 291

QY	1048	CACCCGCGCGTCCAGACAGTGGCCACACAGTGGAGTGGCCGCGCTTC	1107		1461	CTCCACCAAGGCTGGCGCGGTACATGATGATTGGAGCCCTGCCATCAGCAGTTCAGGC	1520	
Db	381	CACGCCGACAGCTAAGAGCAGTACCCACACAGTGTGATGAACCCCAACAGCGCTTT	440		2188	CCTGCGCGGAGCTACAGGCCCTACAGCAGGGTCTCGCGAGGGGGTTCATACCAAC	2247	
QY	1108	GATTGCGCCCTGACAAAGGCAATACCCAGGAACAGTGGAGGCCCGCGCTGCTACTAC	1167		1521	CCTGTGGAGGTTACAGGCCCTATGATAGGGTCTTCGAGGGGGGTGTTATCACCATT	1580	
Db	441	GACTGTGCCCGCCACAAAGGCATACCCAGGAGCAGTGGCGCGCGAGGCTGCTGCTGG	500		2248	GAGACCGGCCAGCCCTGATTGGGAAGGTATGGCCCGGGTCCACTGCCCTTCCCGACTTC	2307	
QY	1168	ATCCCTGCAAGACAGGGCTGCAGGAGCCAGATGGGCGAGCCCTGGTGTCTTCCCA	1227		1581	GAGACTGGCGAGCATTGATTGGGAAGGTTTGGCCCTGGATCCACGCCCTTCCCTGATTTC	1640	
Db	501	GTCCCAAGCAGGAGGTGCTGTAATGGGCCAGTGATGGGCGAGCCCTGGTGTCTTCCCT	560		2308	ACCAACCCCAACAGCCCTGGCTGGTGGAGGACATGGTGGCTGAGTTCATCAGCAGGTG	2367	
QY	1228	CCAGCTACCCAGCTACAACTGGAGAACCTGAGCTCCTCTGAATGGGCTACAGGCC	1287		1641	ACCAACCCCTGAGACCTCGACTGGTGGCAGGACATGGTGTCTGAGTTCATCAGCAGGTG	1700	
Db	561	CCAGCTACCCAGCTACCGCTAGAGAACCTGAGCTCTACAGAGTCAGGGTACACAGCC	620		2368	CCCTTCGAGCGCTCTGGATTGACATGAACAGGCTTCCAACTTCATCAGAGGCTCTGAG	2427	
QY	1288	ACCTGACCGGTACCAACCCGACTTCTTCCCAAGGACATCCTGACCTCGGGCTGGAC	1347		1701	CCCTTCGATGGCATGTGGATCGACATGAACGACGTCGACCTTCATCAGAGGCTCTCAG	1760	
Db	621	ACCTGACCGGACCAAGTCCGACTTCTTTTCCAAAGGATGTGCTGACCTTACAGCTGGAG	680		2428	GACGGCTGCCCAACAATGAGCTGGAGAACCCACCCCTACGTGCCCTGGGGTGGTGGGGG	2487	
QY	1348	GTGATGATGAGACTGAGAACCGCTCCACTTCAAGATCAAGATCCAGCTAACAGGCGC	1407		1761	CAGGGCTGCCCTGACATGAACCTAGAGAACCCCCCTATGTGCCCGGGGTGGTGGCGGG	1820	
Db	681	GTGCTATGAGACTGACAGCGCCCTCCACTTTATGATCAAGATCCCACTAGTAAGCGC	740		2488	ACCTTCAGGGCGGCCACCATCTGTGCTCCAGCCACGAGTTTCTCTCCACACTACACAC	2547	
QY	1408	TACGAGGTGCTTGGAGACCCCGCTGTCCAGACCGGGCACCGTCCCCACTCTACAGC	1467		1821	GCCTTCGAGGGCGGCCACCATCTGTGCTCCAGCCACGAGTTCCTCTCTACACACTACAC	1880	
Db	741	TACGAAGTGCCCTCGAGACCCCGCTGTCTGAGCCAGCGGCATCCCCACTTTACAGT	800		2548	CTGCACACACTCTACGGCTGACCGAGCCATCGCTCCACAGGCGCTGGTGAAGGCT	2607	
QY	1468	GTGGAGTTCTCGAGAGCGCTTCGGGGTGATGCTGCAACGGGAGCTGAGCGCGCGTG	1527		1881	CTCCAAACCTTATCGGCTCTACTGAGCAATGCTCCAGCAGGGCCCTAGTCAAGACT	1940	
Db	801	GTGGAGTTCTCGAGAGCGCTTCGGCGTGATGCTTCGAGGAAGCTTGGTGGCGAGTG	860		2608	CGGGGGACACGCCCAATTTGTGATCTCCGCTCGACTTTGTGGCCACGGCCGATACGCC	2667	
QY	1528	CTGCTCAACAGCAGTGGCGCCCTGTCTTTGGGACACAGTTCTCAGCTGTCCACC	1587		1941	CGGGGAAACAGACCTTTTGTGATCTCCGCTCAAACCTTTGCGGGCCACGGCGGTATGCC	2000	
Db	861	CTGCTCAACACACCGTGGCGCCCTGTCTTCTGCTGACAGTTCTGCTGAGGTCTCCAC	920		2668	GGCCTGGAACGGGGACGTGTGGAGCTCTGGGAGCAGCTCGCTCTCCCTGCCAGAA	2727	
QY	1588	TGCTGCCCTCGCAGTATATCAGAGCCTGCGAGCACTCAGTCCCCTGATGTACAGC	1647		2001	GGTCACTGACAGGGGACGTGTGGAGCTTTGGGAGCATCTTGGGACTCTCTGTGCCAGAA	2060	
Db	921	TCACTGCCCTCCAGCACATGGCAGGCTGGGTGAGCAGCTCAGTCCACTCATGCTCAGC	980		2728	ATCTCTGCAATTTAAACCTGTGGGGTGTCTGTGTGGGGCGGAGCTCTGCGGCTTCTCTG	2787	
QY	1648	ACGAGCTGACAGGATCACCTGTGGACCGGGAACCTTGGGCCACGCCCGGTGGAC	1707		2061	ATCTCTGCAATTTAAACCTGTGGGGTGTCTGTGTGGGGCGGAGCTCTGCGGCTTCTCTG	2120	
Db	981	ACCGAATGACCTGCACTACCCCTCTGGAACCGGGATGTGGCAACCTCGCAAGGTGTCAAC	1040		2788	GGCAACACTCAGAGAGCTGTGTGCTGGAACCCAGCTGGGGGCTTCTACCCCTTC	2847	
QY	1708	CTCTAGGGTCTCACCTTTTACTGCGGCTGGAGGACGGCGGGTGGGACACAGGGGTG	1767		2121	GGAAACACACAGAGAGCTGTGTGCTGGAGCCAGTTCGGGGGCTTCTACCCCTTC	2180	
Db	1041	CTCTATGGGTACACCTTTTACTTGGCACTGGAGGACGGTGGCTTGGCTCAGGGTGC	1100		2848	ATCGGGAACCAACAGCCTGTCTGCTGCCCCCAGGAGCCGTACAGCTTCAGCAGCGC	2907	
QY	1768	TTCCTGCTAAACAGCAATGCCATGGATGTGCTGACCGGAGCCCTGCTTAGCTGG	1827		2181	ATGCGGAACCAACAATGACCTGAATAGCTTGGCTCAGGAGCCGTACAGGTTTACGAGAGC	2240	
Db	1101	TTCCTGCTAAATAGCAATGCCATGGATGTGCTGACCGCCAGCCCGGCTTAACTTGG	1160		2908	GCCACAGAGGCTTACAGGAAGGCCCTCACCTGTGCGCTACGCACTCTCTCCCCACCTCTAC	2967	
QY	1828	AGGTGACAGGTGGATCTCGGATGTCTACATCTCTGGGCCACAGGCCCAAGAGCTG	1887		2241	GCACAGCAGGCCTTACAGGAGGCGCTTCACTTTCGCTATGCTTCTACCTTACCTGTAT	2300	
Db	1161	AGGTCAACAGCGGGAATCTCGGATGTATGTTATGTTTCTTAGGCCACAGGCCCAAGAGTGT	1220		2968	ACACTGTTCCACAGGCCCAACCTGCGGGGGAGACCGTGGCCCGGCCCTCTCTCTCTGAG	3027	
QY	1888	GTGACAGTACCTCGACGTGTGGGATACCCCTTCAATGCGGCATCTAGGGGCTTGGGC	1947		2301	ACACTCTTCCAGCGGCCCAACCGTCAAAAGGGGACACGGTGGCCCGGCCCTCTCTCTGAG	2360	
Db	1221	GTCCAAACATACCTGTGATGTCTGGGATACCCCTTCAAGCTCCGCTAGTGGGGCTTGGC	1280		3028	TTCCCNAGGACTTAGCACCTGAGTGTGACCAACAGCTCTCTGTGGGGGGAGGCCCTG	3087	
QY	1948	TTCACCTGTGCGCTGGGGCTACTCCTCCACCGCTATACCGCGCAGGTGTGGAGAAC	2007		2361	TTCCCTTGAGGATCCAGCACCTGGTGTGTGAGCCGCGCAGCTCTGTGGGGGGCGGCCCTG	2420	
Db	1281	TTCACCTGTGCGCTGGGGCTACTCCTCCACCGCCATCGTCCGACAGGTGTGGAGAAC	1340		3088	CTCATCACCCCAAGTGTCCAGGCCCGGGAAGGCCAGTGAAGTGAAGTCTTCCCTTTGGGC	3147	
QY	2008	ATGACAGGGCCCACTTCCCTGAGGCTCCATGGAACGACCTGACATGAGTCTCC	2067		2421	CTTATCACACCTGTGCTTGAACCTGGGAAAACCTGACGTGACAGGCTACTTCCCAAGGCG	2480	
Db	1341	ATGACAGGACGCACTTCCCTGAGGCTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG	1400		3148	ACATGATGACGCTGTGAGAGCGGTGCCAATAGAGGCCCTTGGCAGCCTCCACCCCACT	3207	
QY	2068	CGAGGAGCTTACGTTCAACAGGATGGCTTCCGGGACCTTCCCGGCCATGTGTGAGGAG	2127		2481	ATGTGATCAACCTTGAGATGTGTGAGAGCCCTTGGTGTGAGAGCCCTTCTCTCTCTCA	2540	
Db	1401	CGAGGAGCTTCACTTCAACAGGAGCGCTTTCGAGACTTCCAGACATGTGTGACGAG	1460		3208	GCAGTCTCCCGTGTGAGCCAGCCA---TCCACAGCAGGGGCGAGTGGGTGACGCTCGCGGC	3264	
QY	2128	CTGCAACAGGGCGCGGCTTACATGATGATGATGATGATGATGATGATGATGATGATGATG	2187					

Db 2541 CCAGCCTCGTCTTTCAGATCTATTGTCACGACGAGGTGCTGAGCTGACACTGGAGCC 2600
QY 3265 CCCCTGGACACCATCAACGTCCACCTCCGGGCTGGGTACATCATCCCCCTCGAGGCGCTT 3324
Db 2601 CCCCTGGATACCATCAACGTGACCTCCAGGCGGGGTACATCATCACTGTCAGGGTCCC 2660
QY 3325 GGCCTCACAAACACAGAGTCCGCCAGACAGCCCATGCGCTGCTGCGCTTGGCCCTAACCAAG 3384
Db 2661 AGCCTCACAAACACAGAGTCCGAAAGACAGCCCATGCGCTGCTGCGCTTAACTGAA 2720
QY 3385 GTGGAGAGCCCGAGGAGCTGTTCTGGACGATGGAGAGCTGGAAGTCTGGAG 3444
Db 2721 AGCGGAGAGCTCTGGGAGCTGTTTGGAGACGCGGAGAGCTCGGGGTCTGGAG 2780
QY 3445 CGAGGGGCTTACACACAGGTCTATCTTCTGCGCCAGGAATAACACGATCGTGAATGAGCTG 3504
Db 2781 CGTGGGCTTACACACAGGTCTGTCACCTTCTCAGCCCAAGTAATACCATTTGAACAGTTA 2840
QY 3505 GTACGTGTGACCAAGTACGAGGAGTGGCTGACCTGACCTGACAGAGTACTGCTCTGGCGGTG 3564
Db 2841 GTGCACGTGACCAAGAGGAGGTGAATCTGCACTGAGGAGGTGACCATCTTGGGAGTG 2900
QY 3565 GGCAGGCGCCCGACAGAGTCTCTCCAGGCTGCTCCAACTGCTCCAACTTCACTACAGC 3624
Db 2901 ACCAGAGTCCCAACCCAGGCTCTTCCAAATGGATCTCTGTCTCAATTTCACTACAGC 2960
QY 3625 CCCGACACCAAGTCTCGACATCTGTGCTGCTGCTGATGGAGAGAGTCTTCTCGTC 3684
Db 2961 CCGGACGACAGAGCTGTCCATCCCTGCTCACTGCTGATGGAGAGGTTTCAATC 3020
QY 3685 AGCTGCT 3691
Db 3021 GATTGCT 3027

RESULT 10
BC010210
LOCUS
DEFINITION Mus musculus glucosidase, alpha, acid, mRNA (cdna clone MGC:11484
IMAGE:3969127), complete cds.
ACCESSION BC010210
VERSION BC010210.1 GI:16307342
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 3504)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altshuler, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diachenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Schetz, T.E., Brannstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Brown, C., Raha, S.S., Loguilland, N.A., Peters, G.J.,
Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodrigues, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smallos, D.B.,
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 3504)
Strausberg, R.
Direct Submission

JOURNAL

REMARK
COMMENT

Submitted (05-JUL-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabbs@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LILNL at: <http://image.llnl.gov>
Series: IRAK plate: 17 Row: b Column: 24
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 6679896.

FEATURES
source

1..3504
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="MGC:11484 IMAGE:3969127"
/tissue_type="Mammary tumor. Metallothionien-TGF alpha
model. 10 month old virgin mouse. Taken by biopsy."
/clone_lib="NCI CGAP_Mami"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1..3504
/gene="Gaa"
/db_xref="LocusID:14387"
/db_xref="MGI:95609"
225..3086
/gene="Gaa"
/codon_start=1
/product="glucosidase, alpha, acid"
/protein_id="AAH10210.1"
/db_xref="GI:16307343"
/db_xref="LocusID:14387"
/db_xref="MGI:95609"

gene

CDS

ORIGIN

Query Match 51.3%; Score 1899.8; DB 10; Length 3504;
Best Local Similarity 82.6%; Pred. No. 1.1e-284;
Matches 2187; Conservative 0; Mismatches 457; Indels 3; Gaps 1;
QY 1048 CACCCCGCGCTCCAGAGCAGTGCACACAGTGGAGCTCCCCCAGACGCGCTTC 1107

Db 435 CAGACTGAACAGCCCAAGACACCCACACAGTGTGATGTGCCCCCAGCAGCGCTTT 494
QY 1108 GATTGGCCCTTGACAAGGCATACACCCAGGAAACAGTGCAGGCGCGCGCTGCTGTCTAC 1167
Db 495 GACTGTGCCCCGACAAAGGCACTCTACAGGAGCAATGAGAGCCCGCGCTGCTGTAT 554
QY 1168 ATCCCTGCAAAAGCAGGGGTGACGAGGAGCCCAAGATGAGGAGCCCTGTGCTTTCCCA 1227
Db 555 GTCCAGCAGGCGAGGTGTGAAGGAGCGCAGATAGAGGACGCCCTGTGTGTTCTTCCCT 614
QY 1228 CCCAGTACCCAGCTTACAGCTGGAGAACCTGAGCTCTCTGAAATGGGCTACACGGCC 1287
Db 615 CCCAGTACCCAGCTTACAGCTGGAGAACCTGAGCTCTACAGAGTCGGGGTACACAGCC 674
QY 1288 ACCCTGACCGTACCAACCCCACTTTCTTCCCAAGACATCTGACCCCTGGGCTGGAC 1347
Db 675 ACCCTGACCGTACCAACCGGCTTCTTCCCAAGAGATGTGTGACCTTACAGCTGGAG 734
QY 1348 GTGATGATGGAGCTGAGAACCGCTTCCACTTCAAGATCAAAAGATCCAGCTAACAGGCGC 1407
Db 735 GTGCTGATGGAGACAGACGCGCTTCCACTTCAAGATCAAAAGATCCCTGTAGTAAGCGC 794
QY 1408 TAGGAGTGGCTTGGAGAGCCCGGTGTCCACAGCGGGCACCGTCCCACTTACAGC 1467
Db 795 TAGCAAGTGGCTTGGAGAGCCCGGTGTCCACAGCGGACCAATCCCACTTTACAGC 854
QY 1468 GTGGAGTGTCCGAGGAGCCCTTGGGGGTGATGCTGACACCGGAGCTGACCGCGCGTG 1527
Db 855 GTGGAAATCTCAGAGGAAACCTTTGAGATGATGCTTGTAGGAACTTGTGGCCGAGTG 914
QY 1528 CTGCTGAACACAGCGGTGGCGCCCTGTGTTTGGGACACAGTTCCTCAGCTGTCCACC 1587
Db 915 TTGCTGAACACAAACCGTGGCGCCCTGTGTTCTGCTGACACAGTTCCTGAGCTGTCCACT 974
QY 1588 TCGCTGCCCTCGCAGTATATACAGGCGCTGCGGAGCACTCAGTCCCTGATGCTCAGC 1647
Db 975 TCCCTGCCCTCCAGACATCATCAGGCGCTGGGGAAACACTCAGGCGCACTCATGTCTCAGC 1034
QY 1648 ACCAGCTGCACAGGATCAACCTGTGGAACCGGGAGCTTGGCGCCACGCGCCGTGCGAAC 1707
Db 1035 ACCGACTGGGCTGATATACCTCTTGGAAACCGGGACACACCACTTCGCAAGGTACCAAC 1094
QY 1708 CTCTACGGGTTCACCTTTCTACTTGGCGCTGGAGGACGCGGGGTGCGCAACAGGGGTG 1767
Db 1095 CTCTACGGGTTCACCTTTCTACTTGGCACTGGAGGACGGTGGCTTGGCTCAGCGTGTG 1154
QY 1768 TTCTCTGCTAAACAGCAATGCCATGGATGCTGCTGACGCGAGCGCTGCCCTTACTGTG 1827
Db 1155 TTCTCTGCTAAACAGCAATGCCATGGATGCTGCTGCAACCCAGCGCCCTTAACCTGG 1214
QY 1828 AGGTGCACAGGTGGGATCTCGATGTCTATATCTTCTGGGCGCAGAGCCCAAGAGCGTG 1887
Db 1215 AGGTCAACGGGCGGATCTTGGATGTATGTGTTCTTAGGCGCCAGAGCCCAAGAGCGTT 1274
QY 1888 GTGACAGCTTACCTGACCTGTGGGATACCGGTTTCACTCCGCGCATCTGGGGCTTGGCG 1947
Db 1275 GTGCAACAATACCTGATGTGTGGGATACCCCTTCACTGCTCCATCTTGGGGCTTCGCG 1334
QY 1948 TTCCACTGTGCGCTGGGGCTACTCTCCACCGCTATCACCGCCAGAGTGTGGAGAAC 2007
Db 1335 TTCCACTGTGCGCTGGGGCTACTCTCCAGCGCCATTTGTCCGCGAGGTATGTGGAGAAC 1394
QY 2008 ATGACAGGGGCCACTTCCCTTGGACGTCCCAATGGAACGACCTGGAATACATGGAATCC 2067
Db 1395 ATGACCAAGACACACTTCCGCTGGATGTGCAATGGAATGCACTGCACTACATGGAAGCC 1454
QY 2068 CGAGGGAATTCAGTTCATCAAGAGATGGTTCGGGATCTTCCGGCATGTGTCAGGAG 2127
Db 1455 CGAAGAGACTTCACTTCAACAGGACAGCTTTGCCGATCTTCCAGACATGTTGCGGGAG 1514
QY 2128 CTGCAACGGGCGCGGCTTACATGATGATGCTGGATCTGCTGCATCAGCAGCTCGGCG 2187
Db 1515 CTGCAACGGGTTGGCGGCTTACATGATGATGCTGGATCCCTGCTCATCAGCAGCGCAGGC 1574

QY 2188 CTTGCGGGAGCTACAGGCCCTTACGACGAGGGTCTGCGAGGGGGTTTTTCATCACCAC 2247
Db 1575 CTTGCTGGAGTTACAGGCCCTTACGACGAGGGTCTGCGAGGGGTGTGTTTCATCACCAC 1634
QY 2248 GAGACGGGCGACCGCTGATTTGGGAAAGGTATGGCCCGGGTCCACTGCGCTTCCCGACATTC 2307
Db 1635 GAGACTGGGCGACCGCTGATTTGGGAAAGGTATTTGGCCCGGAAACACCGCTTCCCTGATTTTC 1694
QY 2308 ACCAACCCACACAGCCCTGCGCTGTGGGAGGACATGTGGCTGAGTTTCACTGACAGGCTG 2367
Db 1695 ACCAACCTTGAGACCCCTTGAATGTGTGGCAGGACATGTGTCTGAGTTCCACGCCCAGGTTG 1754
QY 2368 CCTTTGCAACGGCTTGTGATTTGACATGAACGAGCCCTTCCAACTTTCATCAGAGGCTCTGAG 2427
Db 1755 CCTTTGATGGCATGTGGCTGCAATGAACGAAACCGTCCAACTTTCGTTAGAGGCTCTCAG 1814
QY 2428 GAGGCTGCCCCCAACATGAGCTGAGGAGAACCCACCTTACGTCCTGGGGTGTGGGGG 2487
Db 1815 CAGGGCTGCCCCCAACATGAACCTGGAGAACCCCCCTTATGTGCCCGGGGTGTGGCGGG 1874
QY 2488 ACCCTCCAGGGGCGACCATCTGTGCTCCAGCCACCACTTCTCTCCACACACTACAAAC 2547
Db 1875 ATCTTGCAGGCGAGCCCACTGTGTGCTTCCAGCCACCAATTCCTCTCCACACACTACAAAC 1934
QY 2548 CTGCAACACCTTACGGCTTGAACGAAAGCCATGCGCTCCCAACAGGCGCTGTGAAGGCT 2607
Db 1935 CTCCACAACTTGTACGGCTTCACTGAAGCTATCGGCTCCAGCAGGGCGCTGTGTAAGACT 1994
QY 2608 CGGGGAGACGCGCCATTTGTGATCTCCGCTCGACCTTTGTCTGGGCAACGGCCGATACGCC 2667
Db 1995 CGGGGAAACGACCCCTTTGTGATCTTCCGCTCAACCTTCTCGGGGCAACGGCCGCTACGCT 2054
QY 2668 GSCCACTGACCGGGGAGCGTGGAGCTCTGGGAGCAGCTCGCTCTCTCCGTCGACAGAA 2727
Db 2055 GGTCACTGACAGGGGATGTGGGAGCTCTTGGGAGCATCTTGGCATCTCTCTGTGCCAGAC 2114
QY 2728 ATCTCTGCAATTAACCTCTGCGGGTGTCTTGTGTGGGGCGGACGCTGTGCGGCTTCCTG 2787
Db 2115 ATCTCTGCAATTAACCTCTGCGGGTGTCTTGTGTGGGGCGGACATCTGCGGCTTCATA 2174
QY 2788 GSCACACCTCAGAGAGCTGTGTGCTGGACCCAGCTGGGGGCTTCTTACCCCTTC 2847
Db 2175 GGAGACAGCTCAGAAAGCTGTGTGCTGGACCCAGCTTGGGGGCTTCTTACCCCTTC 2234
QY 2848 ATGCGGAAACCAACAGCCTCTCAGTCTGCGCCAGGAGCCGTACAGCTTTCAGCGAGCG 2907
Db 2235 ATGCGGAAACCAACATGACCTGATAGCGTCTCTGAGGAGCCGTACAGGTTTCAGCGAGAG 2294
QY 2908 GCCCAGAGGCGCATGAGAAAGCCCTTCACTGCGCTAGCAGCTCTCTGCGGAGGAGGCGCTTC 2967
Db 2295 GCGCAGAGGCGCATGAGAAAGGCTTTCGCTTACGCTATGCGCTTCTGCGCTTACCTGTAC 2354
QY 2968 ACACCTGTTCCACAGGCGCCAGCTGCGGGGGAGACCGTGGCCCGGCGCTTCTTCTGGAG 3027
Db 2355 ACTCTCTTCCACCGCGCCACCGTCAAGAGAGACACGGTGGCGGCGCTTCTTCTGGAG 2414
QY 3028 TTCCCCAAAGGACTCTAGCACCTTGGAGCTGTGGACCAACAGCTCTCTGTGGGGGAGGCGCTTC 3087
Db 2415 TTCCCTTGAGGATCCAGGACACTGTGTCTGTGACCCGCGAGCTCTTGTGGGGGCGGCGCTTC 2474
QY 3088 CTTCATCAACCCAGTCTCCAGGCGGGGAGGCGGAGGTGACTGGGTACTTTCCCTCTTGGCG 3147
Db 2475 CTTCATCAACCTTGTGAGCTTGGAGCTGGGAAACTGAAGTGAACGGGCTTCTTCCCAAGGCG 2534
QY 3148 ACATGGTACGACTCGACAGCGGTGCAATAGAGGCGCTTGGAGGCTTCCACCCCACTT 3207
Db 2535 ACGTGGTCAACATGACAGGTGTGTCAGTGGATTCCCTCGGTACTTCTCCCTTCTCCATCA 2594
QY 3208 GCAGCT---CCCGGTGAGCCAGCCATCCACAGCAGGGGCGAGTGGGTGACGCTGCGCGGCG 3264
Db 2595 TCGGCTTTCATCTTTCAGATCTGCTCTCCAGACAGGGGCGAGTGGCTGACACTGGAGGCC 2654

```
QY 3265 CCCTTGGACACCATCAACGTCCACCTCCGGGCTGGGTACATCATCCCTCGAGGGCCCT 3324
Db 2655 CCACTGGATACCATCAACGTGCACTTGGAGGGGTATCATATACCGCTCGAGGGTCCC 2714
QY 3325 GGCTTCAACACACAGAGTCCGCGCAGAGCCCATGGCCCTGCTGGCCCTTAACCAAG 3384
Db 2715 AGCTTCAACACACAGAGTCCGCGCAGAGCCCATGGCTCTGGCTGGGTATTAACAGCA 2774
QY 3385 GGTGGAGAGCCCGAGGGAGCTGTTCTGGGACGATGGAGAGCGCTGGAAAGTGTGGAG 3444
Db 2775 AGCGGCGAGCGGATGGGAGCTGTTCTGGGACGAGCGGGAGAGCGCTTGCAAGTCTGGAG 2834
QY 3445 CGAGGGCCCTACACACAGGTTCATCTTCTGGGCCAGGAATAACACGATCGTGAATGCTG 3504
Db 2835 CGTGGGGCTTACACATGTGTCACTTCTCAGCGCAAGAACAAATACCATTTGAACAAGTTA 2894
QY 3505 GTACGTGTACACAGTGAAGGAGCTGGCTGACGCTCAGAGGTGACTGTCTGGGCGGTG 3564
Db 2895 GTGCGTGTACCAAGAGAGGAGCTGAGCTTACACTAGGAGGTGACCGTCTTGGAGGTG 2954
QY 3565 GCCACGGCCCGCAGAGGTCTCTTCCAAAGGTGTCCTGTCTCCAACTTCACTTACAGC 3624
Db 2955 GCCACAGCTCTTACCAGGTCTTCTCCAAAGGATCCCTGTCTCCAACTTCACTTACAGC 3014
QY 3625 CCGACACCAAGTCTCTGACATCTGTGTCTGCTGTGATGGAGAGAGTTTCTCGTC 3684
Db 3015 CTTGACAAAGAGCTGGCCATCTCTGTCTCACTGCTGATGGAGAGCTGTTTCAATC 3074
QY 3685 AGCTGCT 3691
Db 3075 AGCTGCT 3081

RESULT 11
AX305670
LOCUS AX305670 3364 bp DNA linear PAT 11-DEC-2001
DEFINITION Sequence 421 from Patent WO0188188.
ACCESSION AX305670
VERSION AX305670.1 GI:17645110
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1
AUTHORS Ishikawa,K., Arai,S., Takahashi,Y., Nagata,T. and Ishii,Y.
TITLE Method for examining ischemic conditions
JOURNAL Patent: WO 0188188-A 421 22-NOV-2001;
School Juridical Person Nihon University (JP)
FEATURES
source
1..3364
/organism="Mus musculus"
/mol_type="unassigned DNA"
/db_xref="taxon:10090"

ORIGIN
Query Match 50.8%; Score 1881.4; DB 6; Length 3364;
Best Local Similarity 82.1%; Pred. No. 8.2e-282;
Matches 2174; Conservative 2; Mismatches 468; Indels 3; Gaps 1;

QY 1048 CACCCCGGCGTCCACAGAGAGTCCACACAGTGGCGAGTCCCGCCCAACAGCGCTTC 1107
Db 323 CAGACTGAACAGCCCAAGAACACACACAGTGTGATGTGCCCCCAGCAGCGCTTT 382
QY 1108 GATTGGCCCTTACAGAGCCATCAACAGGACAGTGGAGGCCCGCGGCTGCTCTAC 1167
Db 383 GACTGTGCCCGCCCAAGAGCATCTCAGAGGACAAATGCGAGGCCCGAGCTGCTCTA 442
QY 1168 ATCCCTGCAAGAGGAGGCTGAGGAGAGCCAGATGGGAGCGCTTCTTCCCA 1227
Db 443 GTCCAGAGGGGAGGTGCTGAAGAGGCCACAGATAGGGCAGCCCTTCTTCTCCCT 502
QY 1228 CCAGCTACCCAGCTACAGCTGGAGAACCTGAGCTCCTCTGAAATGGGCTACAGGCC 1287
```

```
Db 503 CCAGCTACCCAAAGTACCGTCTAGAGAACCTGAGCTCTACAGAGTCGGGGTACACAGCC 562
QY 1288 ACCCTGACCCGTACCAACCCCACTTTCTTCCCAAGGACATCTTACACCTCGGGCTGGAC 1347
Db 563 ACCCTGACCCGTACCAACCCCACTTTCTTCCCAAGGATGTCTGACCTTACAGCTGGAG 622
QY 1348 GTGATGATGGAGACTGAGAAACCGCTCTCACTTCAAGATCAAAAGATCCAGCTAACAGCGC 1407
Db 623 GTGCTGATGGAGACAGACAGCGCCCTCCACTTCAAGATCAAAAGATCCCAAGTAAAGCGC 682
QY 1408 TACGAGGTGCTTGGAGACCCCGGTGTCACAGCGGGCACCGTCCCACTCTACAGC 1467
Db 683 TACGAGGTGCTTGGAGACCCCACTGCTGTGAGCCAGGACCACTCCCACTTACAGC 742
QY 1468 GTGAGATTCTCGAGGAGCCCTTCCGGGTGATCTGTACACCGGAGCTTGAACGCGCGTG 1527
Db 743 GTGGAATTCTCAGAGGAACCCCTTGGAGTGTCTGTAGGAAGCTTGTGTGCGAGTG 802
QY 1528 CTGCTGAACACAGCGGTGGCGCCCTTCTTGGGACCACTTCTTCACTGCTGTCCACC 1587
Db 803 TTGCTGAACACAGCGGTGGCGCCCTTCTTGGGACCACTTCTTCACTGCTGTCCACT 862
QY 1588 TCGCTGCCCTCGCAGTATATACAGCGCTCGCGGAGCACTCAGTCCCTGATGCTCAGC 1647
Db 863 TCCCTGCCCGCCGACATACAGCGCTTGGGGAAACCTCAGCCCACTCATGCTCAGC 922
QY 1648 ACCAGCTGACACAGGATCACCCCTGTGGAAACCGGACCTTTCGCCCAACAGCTTGGAAAC 1707
Db 923 ACCGACTGGGCTGTATCACCTCTGGAACCGGACACACACCCCTCGCAAGTACCAAC 982
QY 1708 CTCTACGGGTCTCACCCCTTCTTACCTGGCGTGGAGAGCGCGGTGCGCAACAGGGGTG 1767
Db 983 CTCTACGGGTCTCATCTTCTTACCTGGCACTGGAGAGCGGTGCTGCTCACGGTGC 1042
QY 1768 TTCCTGCTTAAACAGCAATGCCATGATGTGGTCTGTCAGCGGAGCCCTGCTTACGCTG 1827
Db 1043 TTCCTGCTTAAACAGCAATGCCATGATGTATCTTCTAGGCCAGAGCCCAAGAGCGTC 1162
QY 1828 AGGTGACAGGTGGGATCTGATGTCTACATCTTCTTGGGCCAGAGCCCAAGAGCGTG 1887
Db 1103 AGGTCAACGGGCGGATCTTGGATGTGTATGTCTTAGGCCAGAGCCCAAGAGCGTC 1162
QY 1888 GTGACAGAGTACTGGAAGTGTGGATACCGTGTATGCTCCGCCATATCTGGGGCTGGGC 1947
Db 1163 GTGCAACAATACCTGGATGTGTGGGATACCCCTTATGCTCCATCTTCTGGGCTCGGC 1222
QY 1948 TTCACCTGTGCGCTGGGCTACTCTCCACCGCTATCACCGCCAGGTGGTGGAGAAC 2007
Db 1223 TTCACCTGTGCGCTGGGCTACTCTCTCGACCGGCAATGTCGCGCAGGTAGTGGAGAAC 1282
QY 2008 ATGACACAGGGCCCACTTCCCGCTGGAAGTGTGGAAGGAGTGTGGAAGTGTGGAAGTGT 2067
Db 1283 ATGACACAGGACACATCTCCCGCTGGAAGTGTGGAAGTGTGGAAGTGTGGAAGTGT 1342
QY 2068 CGGAGGAGTCTTCAAGTGTGGAAGTGTGGAAGTGTGGAAGTGTGGAAGTGTGGAAGTGT 2127
Db 1343 CGAAGAGACTTCACTTCAACCCAGGACAGCTTTCGCGCATCTCCAGACATGCTGGCGGAC 1402
QY 2128 CTGACACAGGGCGCGGCTTACATGATGTGGAAGTGTGGAAGTGTGGAAGTGTGGAAGTGT 2187
Db 1403 GTGCAACAGGGTGGCGGCTTACATGATGTGGAAGTGTGGAAGTGTGGAAGTGTGGAAGTGT 1462
QY 2188 CTGCGCGGAGCTACAGGCTTACAGGAGGTGTGCGGAGGGGTGTTTCAATCACCAAC 2247
Db 1463 CTGCTGGAGTACAGGCTTACAGGAGGTGTGCGGAGGGGTGTTTCAATCACCAAC 1522
QY 2248 GAGACCGGCGAGCTGATGGGAGGTGTGCGGAGGTGTGCGGAGGTGTGCGGAGGTGTGCGGAGGTGT 2307
Db 1523 GAGACTGGGCGAGCTGATGGGAGGTGTGCGGAGGTGTGCGGAGGTGTGCGGAGGTGTGCGGAGGTGT 1582
QY 2308 ACCAACCCACAGCGCTTGGCGTGTGGGAGGACATGTTGGTGTGTTCCATGACCAAGGTG 2367
```


ELPDDGESLAVLEHGAAYTLVTFSAKNWTVNKLVRVTKGAEHLQLKEVTVLGVATAP
TQVLSNGIPVSNFTYSPDNKSLAIPVSLMLGELFQISWS*

ORIGIN

Query Match 50.8%; Score 1881.4; DB 10; Length 3364;
Best Local Similarity 82.1%; Pred. No. 8.2e-282;
Matches 2174; Conservative 2; Mismatches 468; Indels 3; Gaps 1;

1048 CACCCGGCGGTGCCAGAGCAGTGCACACACAGTGCAGGTCCCCCCCCAACAGCCGCTTC 1107
1108 GATTGGCCCGCTGACAAGGCATCACCCAGAGAACAGTGCAGGCCCGCGGCTGCTGTAC 1167
383 GACTGTGCCCGCCGACAAAGGCATCTCACAGGAGCAATGCGAGGCCGAGGCTGCTGTAT 442
1168 ATCCCTGCAAGCAGGGGTGCAGGAGGCCAGATGGGCGAGCCCTGTGTCTTCCCA 1227
443 GTCCCAAGCAGGCGAGTGTCTGAAGGAGCCACAGATAGGGAGCCCTGTGTCTTCCCT 502
1228 CCAGCTACCCCAAGCTACAAAGCTGGAGAACCTTGAGCTCCTCTGAAATGGGCTACACGGCC 1287
503 CCAGCTACCCCAAGCTACCGCTCTAGAGAACCTTGAGCTCTACAGAGTCGGGGTACACAGCC 562
1288 ACCCTGACCGGTACACCCCGACCTTCTCCCAAGGACATCTGACCCCTGGGCTGGAC 1347
563 ACCCTGACCGGTACACCGCGGACCTTCTCCCAAGGATGTGTGACCTTACAGCTGGAG 622
1348 GTGATGATGGAGCTGAGAACCGCTCCACTTCCAGATCAAGATCCAGCTAACAGCGGC 1407
623 GTGCTGATGGAGACAGACAGCGCCCTCCACTTCAAGATCAAGATCCCGCAAGTAAGCGC 682
1408 TAGAGGTGCTCTGGAGACCCCGCGGTCTCCAGACCGGGGACCGTCCCGCACTCTACAGC 1467
683 TAGGAAGTGCCTTGAGACCCCAAGCTGTGCTGAGCAGGACCACTCCCGCACTTACAGC 742
1468 GTGAGTCTTCGAGAGGCGCTTTCGGGTGATGCTGCAACGGGAGCTGAGACGGCGCGTG 1527
743 GTGGAATCTCAGAGGAAACCTTTGGAGTGTGCTTCTGAGGAGCTTGGTGGCCGAGTG 802
1528 CTGCTGAAACAGAGGTGGGCGGCTTCTTGGGACAGTTCCTTCAAGCTGCCAC 1587
803 TTGCTGAAACAAACCGTGGCGGCGGCTTCTTCTGCTGACAGTTCCTGAGCTGTCCA 862
1588 TGCTGCTGCTCGAGTATACAGAGCTCGCGGACACCTAGTCCCTGATGCTCAGC 1647
863 TCCTGCGCGCCAGACATACAGAGCTGGGGGACACCTCAGCCCACTCATGCTCAGC 922
1648 ACCAGCTGGAACAGGATCAACCTGTGGAAACCGGACCTTTCGCGCCACGCCGCTGCGAAC 1707
923 ACCGACTGGGCTCGTATACCTCTGGAAACCGGACACACACCTCGCAAGGTACCAAC 982
1708 CTCACGGGTCTCACCTTCTACCTGGCGCTGGAGGACGGGGTTCGGCACAGGGGTG 1767
983 CTCACGGGTCTCATCTTCTTACCTGGCACTGGAGGACGGTGGCTTGGCTACGGTGT 1042
1768 TTCCTCTTAAACAGCAATGCATGAGTGTGCTTGCAGCCGAGCCCTGCGCTTACCTG 1827
1043 TTCCTGCTTAAACAGCAATGCATGAGTGTATCTCTGCAACCCAGCCAGCCCTTAACTGG 1102
1828 AGGTGCACAGGTGGGATCTCTGATGTCTATCTCTCTGGGCCAGAGCCCAAGAGCGTG 1887
1103 AGGTCAACGGGCGGGATCTCTGATGTGTGTTCTTAGGCCACAGAGCCCAAGAGCGTC 1162
1888 GTGACAGGATCTCTGAGCTGTGGGATACCCGTTTCAATGCGCGCATCTGGGGCTGGGC 1947
1163 GTGCAACAAATACCTGATGTGTGGGATACCCCTTTATGCTCTCACTTGGGGCTCGGC 1222
1948 TTCACCTGTGGCGCTGGGCTACTCTCTCAAGCTATCACCGCCAGAGGTGTGGAGAAC 2007
1223 TTCACCTGTGGCGCTGGGCTACTCTCTGACCGCAATGTCTGCGCAGAGTGTGGAGAAC 1282
2008 ATGACAGGGGCCCACTTCCCTCGGAGCTCAATGGAAACGACCTGGACTACATGGACTCC 2067

1283 ATGACAGAGACACACTTCCCGCTGACGTGCATGAAATGAACTGACTGTGACTATGACGCGC 1342
2068 CCGAGGGAGCTTCACTTCAAAAGGATGCTTCCGGGACTTCCCGGCCATGATGTCAGGAG 2127
1343 CGAAGAGACTTCACTTCAAAAGGATGCTTCCGGGACTTCCCGGCCATGATGTCAGGAG 1402
2128 CTGCAACAGGGGGCGGCGCTTACATGATGATGTCGTCGTCCTCATGACAGCTGCGGC 2187
1403 GTGCAACAGGGGTGGCGGCTTACATGATGATGTCGTCGTCCTCATGACAGCTGCGGC 1462
2188 CCGCGGGAGGCTACAGGCGCTTACAGAGGCTTCCGGAGGGGGTTCATACACCAAC 2247
1463 CCGTGGAGGTTACAGGCGCTTACAGAGGCTTCCGGAGGGGGTTCATACACCAAC 1522
2248 GAGACCGGCGGCGCTGATTTGGGAAGGTATGGCGCGGCTCCACTGCTTCCCGCACTTC 2307
1523 GAGACTGGGCGGCTGATTTGGGAAGGTATGGCGCGGCTCCACTGCTTCCCGCACTTC 1582
2308 ATCAACCCCAAGCGCTTGGGAGGAGCATGTCGTCGTCGTCCTCATGACAGCTGTCAG 2367
1583 ACCAAACCGCTGAGAGCGCTTACAGGCTTGGGAGGAGCATGTCGTCGTCGTCCTCAT 1642
2368 CCCTTGGAGCGCTTGGATGACATGACAGGCTTCCCACTTCCATGACAGCTGTCAG 2427
1643 CCCTTGGAGCGCTTGGATGACATGACAGGCTTCCCACTTCCATGACAGCTGTCAG 1702
2428 GACGCTGCGGCGGCTGAGTGGAGAACCCCACTTACGTCGCTGGGCTGGTGGGGG 2487
1703 CAGGCTGCGGCGGCTGAGTGGAGAACCCCACTTACGTCGCTGGGCTGGTGGGGG 1762
2488 ACCCTTGGAGCGGCTGAGTGGAGAACCCCACTTACGTCGCTGGGCTGGTGGGGG 2547
1763 ATCTTGGAGCGGCTGAGTGGAGAACCCCACTTACGTCGCTGGGCTGGTGGGGG 1822
2548 CTGCAACAGCTTACGCGCTGACCGAGGCTTCCCACTTCCAGGCGCTGGTGGAGGCT 2607
1823 CTCCAACAGCTTACGCGCTGACCGAGGCTTCCCACTTCCAGGCGCTGGTGGAGGCT 1882
2608 CCGGCGGACCGGCTTACGCGCTGACCGAGGCTTCCCACTTCCAGGCGCTGGTGGAGGCT 2667
1883 CCGGCGGACCGGCTTACGCGCTGACCGAGGCTTCCCACTTCCAGGCGCTGGTGGAGGCT 1942
2668 GCGCACTGGAACGGGCGGCTGAGGCTTCCGGAGGAGCTGCGCTTCTCCGTCGAGAA 2727
1943 GGTCACTGGAACGGGCGGCTGAGGCTTCCGGAGGAGCTGCGCTTCTCCGTCGAGAA 2002
2728 ATCTGCAAGTTAACTGCTGGGCGGCTTCCGGAGGAGCTGCGCTTCTCCGTCGAGAA 2787
2003 ATCTGCAAGTTAACTGCTGGGCGGCTTCCGGAGGAGCTGCGCTTCTCCGTCGAGAA 2062
2788 GCGCAACAGCTTACGCGCTGAGGAGGCTTCCGGAGGAGCTGCGCTTCTCCGTCGAGAA 2847
2063 GCGCAACAGCTTACGCGCTGAGGAGGCTTCCGGAGGAGCTGCGCTTCTCCGTCGAGAA 2122
2848 ATGCGGAACCAACAGGCTTCTGAGTTCGCGGAGGAGGCTTACAGCTTACAGGAGGCT 2907
2123 ATGCGGAACCAACAGGCTTCTGAGTTCGCGGAGGAGGCTTACAGCTTACAGGAGGCT 2182
2908 GCGCAACAGGCTTACAGGAGGCTTCCAGCTTCCGGAGGAGGCTTCCAGCTTCCAGCTTAC 2967
2183 GCGCAACAGGCTTACAGGAGGCTTCCAGCTTCCGGAGGAGGCTTCCAGCTTCCAGCTTAC 2242
2968 ACACCTTCCAGGAGGCTTCCAGGAGGCTTCCAGCTTCCGGAGGAGGCTTCCAGCTTAC 3027
2243 ATCTTCTCCAGGAGGCTTCCAGGAGGCTTCCAGCTTCCGGAGGAGGCTTCCAGCTTAC 2302
3028 TTCCCGCAAGGAGCTTACAGCTTGGAGCTTGGAGCTTGGAGCTTGGAGGAGGCTTGG 3087
2303 TTCCCGTGGAGCTTCCAGGAGCTTGGAGCTTGGAGCTTGGAGCTTGGAGGAGGCTTGG 2362
3088 CTCATCACCGGAGGCTTCCAGGAGGCTTGGAGCTTGGAGCTTGGAGCTTGGAGGAGGCTTGG 3147

Db	2363	CTCATCACACCTGTGCTTGAGCGCTGGAAACTGAAGTACGCGGTACTTCTCCCAAGGCG	2422
QY	3148	ACATGTGTACGACCTGCAGACGGTGCACATAGAGGCGCTTGGAGCGCTCCACCCCACT	3207
Db	2423	ACGTGTGTAACATGACGGTGTGTCAGTGGATTCCTCGGTACTCTCCTTCTCATCA	2482
QY	3208	GCAGCT---CCCCGTGAGCCAGCCATCCACAGCGAGGGCAGTGGGTGACGCTGCCGGCC	3264
Db	2483	TGGCTTTCATCTTCCATGCTGCTCCAGACGACGAGGGCAGTGGCTGACACTGGAAGCC	2542
QY	3265	CCCTGACACCATCAAGCTCACTCCGGGTGGGTATCATCATCCCTGAGGGCCCT	3324
Db	2543	CCACTGGATACCATCAACGCTGACCTGAGGGAGGGGTATCATCATCCGCTGAGGGTCCC	2602
QY	3325	GGCTCACAACACACAGTCCCGCCAGCAGCCATCGGCTGCTGGCCCTTAACAAG	3384
Db	2603	AGCTCACAACACAGAGTCCCGAAGACAGCCCATGGCTCTGGCTGTGGCATTAACAGCA	2662
QY	3385	GGTGGAGAGGCCCGAGGGAGCTGTTCTGGGACGATGGAGAGCGCTGGAAGTCTGGAG	3444
Db	2663	AGCGCGAGCGCGATGGGAGCTGTTCTGGACGACGGGGAGACCTTCGAGTCTTGGAG	2722
QY	3445	CGAGGGGCTACACACAGTCACTTCTTGGCCAGCAATTAACAGTCTGATGAGCTG	3504
Db	2723	CATGGGGCTTACACACTGTGTCACTTCTCAGCCCAAGAACATACCATTTGGAACATTA	2782
QY	3505	GTACGTGTGACACAGTCAAGGAGCTGCTTCCACGCTGACAGAGTGACTGCTGGCGGTG	3564
Db	2783	GTGCGTGTGACAGAGGAGGAGCTGAGCTTAACTGAAGAGTGACCTCTTGGAGTG	2842
QY	3565	GCCAGCGGCCCGCAGAGTCTCTCCACGCTGCTGCTGCTTCCAACTTCACTACAGC	3624
Db	2843	GCCACAGCTCTTACCCAGGTCTTCTCCAGCGCATCCCTGTCTCAATTTCACTACAGC	2902
QY	3625	CCGACACACAGTCTCTGACATCTGTCTGTCTGTCTGATGGAGAGCAGTTTCTCGTC	3684
Db	2903	CTTGACAAACAGAGCGCTGCGCATCCCTGTCTCACTGCTGATGGAGAGCTGTTCAAATC	2962
QY	3685	AGCTGCT 3691	
Db	2963	AGCTGCT 2969	
RESULT 13			
AB006754			
LOCUS			
DEFINITION			
AB006754 3112 bp mRNA linear VRT 16-FEB-2002			
Coturnix japonica GAAII mRNA for acid alpha glucosidase, complete			
cds.			
ACCESSION			
AB006754			
VERSION			
AB006754.2 GI:18693230			
KEYWORDS			
SOURCE			
ORGANISM			
Coturnix japonica (Japanese quail)			
Coturnix japonica			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;			
Phasianinae; Coturnix.			
REFERENCE			
1			
Kunita, R., Nakabayashi, O., Wu, J. Y., Hagiwara, Y., Mizutani, M.,			
Pennybacker, M., Chen, Y. T., and Kikuchi, T.			
Molecular cloning of acid alpha-glucosidase cDNA of Japanese quail			
(Coturnix coturnix japonica) and the lack of its mRNA in acid			
maltese deficient quails			
Biochim. Biophys. Acta 1362 (2-3), 269-278 (1997)			
JOURNAL			
MEDLINE			
PUBMED			
9540858			
2 (bases 1 to 3112)			
Kikuchi, T.			
Direct Submission			
Submitted (20-AUG-1997) Takeki Kikuchi, National Institute of			
Neuroscience, NCNP, Department of Animal Models for Human Disease;			
4-1-1 Osawahigashimachi, Kodaira, Tokyo 187, Japan			
[E-mail:tkikuchi@ncnp.nippon.ac.jp, Tel:0423-41-2711,			
Fax:0423-46-1754]			

COMMENT	On Feb 15, 2002 this sequence version replaced gi:3097299.		
FEATURES	Location/Qualifiers		
source	1..3112		
	/organism="Coturnix japonica"		
	/mol_type="mRNA"		
	/db_xref="taxon:93934"		
	/note="synonym: Coturnix coturnix japonica"		
gene	1..3112		
	/gene="GAAII"		
CDS	7..2628		
	/gene="GAAII"		
	/codon_start=1		
	/product="acid alpha glucosidase"		
	/protein_id="BAA25890.2"		
	/db_xref="GI:18693231"		
	/translation="MALILLRLSPSASSADSACSLPPDERDCGPERLLARADC EARGCYAPSSGSGGPPWCFPLGYRSYRADNVNTATAGGYSARLRVVPFLPADV GTLRDVAMETESRLRTPRPAKQRYEVPMTAPRVSPTAADTLTGVLQDPPFGIV FRQPDQVLLNTVAPLFFADQLQISTSLPSRFISGGERLAPLILDTATMTKTLWN RDMAPQVNLVYSHPFYVLEDDGSAHGFLNNSAMDVLQPCALPTWRTTGGILD FVILFGPDQSVVOQYLDVVGPPFMPYALGFHLRCRWGYSVSTTRQAAANMAGLF PLDVQNDLDYMDAKRFTYKNETPRDYDMVHDFHORGLHVMIVDPGISSSGPGT YRPVDDGLKRGVFIENATQQLIKVWPGTAPDPTNPETHWHDHVMKQPEQVVF DMDIDMNEPSNFVEGSDQDPSLSLEKPPYVPGVGGRLQAGTICASSQQVLSHYN LHSLYGLTEATASHNALLRVGRTRPFVISRSTFAGHYAGHWTGDVSSWQLSAHV PEVLINLLGLVPLGADICGAGDTSEELCVRTQLTGYFPFMRNHNDRNRPQBYA FSLPAQDAMRSALRLRYLSLPVLYTLFRAHMAGDTVARPLPLEPKDPNTWIDRQL LWAGLLTTPVLEQGTQVSGVFPAGTWYSFTGDSITHSKQWILAAALPDTINVHIR AGHLPLEQRLNTAESRKGMVVALTPDGFARGELPFDGDSWQSFKEGDCETIL FLIARGAVLSQLRAGHGDLILVEATVLGVPSAPQRLVLANGIPVEDFSYRSDDTVL RVSVSLPMWEQFVVPWS"		
	polyA_site 3082		
	/gene="GAAII"		
	/note="19 A nucleotides"		
ORIGIN			
Query Match	36.0%; Score 1332; DB 5; Length 3112;		
Best Local Similarity	69.9%; Pred. No. 1.1e-196;		
Matches 1856; Conservative	0; Mismatches 740; Indels 60; Gaps 2;		
QY	1041	CGGTGCACACCCCGCGCTCCAGAGCAGTCCCAACAGTCCGACGCTCCCGCCCAACAG	1100
Db	33	CGGCTCGTGGCCCTCGCTCCAGCTCCGCGGACTCCGCTTTCGCTGCCCGCGGACGA	92
QY	1101	CGCTTCGATTGCGCCCTCGACAGGCCATACCCAGGAACAGTTCGAGGCCCGCGGCTG	1160
Db	93	AGGATTCGACTCGGCGCCCGAGCGCTGCTGGCGGGCGGAGCTGCGAGCGCGGGCTG	152
QY	1161	CTGTACATCCCTGCAAGACGAGGGCTGACGAGGAGCCACAGATGGGGGAGCCCTGTGCTTT	1220
Db	153	CTGTACGCTCGCTCGGCTCGGCGGCTCGGCG-----GGGCTCCCTGCTGCTT	203
QY	1221	CTTCCCAACCGAGTACCCAGCTACAGCTGGAGAACCTGAGTCTCTGAAATGGGCTA	1280
Db	204	CTTTCCTCTCGGCTACCCAGCTACAGGCGGAGCAACGTGACGCGCAACCGCGCGGATA	263
QY	1281	CACGCGCACCTTGACCCGCTACACCCCACTTCTTCCCAAGGACATCTGACCCCTGCG	1340
Db	264	CTCGGCGAGGCTCCGCGCGGTGTCACAGTCTCTGCGCGGAGCTGGGACCGTGGCG	323
QY	1341	GCTGACGTGATGATGGAGACTGAGAACCGCTCCATCTTACGATCAAGAATCCAGCTAA	1400
Db	324	GTTCGACGTGGCGATGGAGACCGAGAGCGCTTGGGCTTTCAGCCCTCGCGATCCGCG	383
QY	1401	CAGGCGCTACGAGGTGCCCTTGGAGACCCCGCTGTCACAGCGGGGACCGTCCCCACT	1460
Db	384	GCAGCGCTACGAGGTGCCCATGCGCCACCGAGGCTGAGCACCCGCGGCGAGACACGCT	443
QY	1461	CTACGCGTGAAGTCTCCGAGGAGCCCTTCCGGGTGATCTGTGTCACCGCGAGCTGAGCGG	1520
Db	444	GTACGCGGTGACGCTCTCCAGGATCCGTTCCGATCGTGGTTCGCGACCCCGATGG	503
QY	1521	CCGCGTGTCTGTGAACAGACGAGTGGCGCCCTGTCTTTTGGGAGCCAGTCTCTTCAGCT	1580

504	GCAGGTCCTGCTCAACACCAACGCTCGCTCCGCTGTTCTTTTCGCGACAGGATTCCTCGAGAT	563
1581	GTTCCACACTCGCTGCCTCGCAGTATATACAGAGCCTCGCCAGCACCTCAGTCCCCTGAT	1640
564	CTTCACCTCCCTGCTTCCGTTTATTTCTGGGCTGGGGAGCGTCTGGCCCTCTCTCAT	623
1641	GCTCAGCACCAAGCTGGACCAAGGATCACCCCTGTGGAAACCGGAGCCTTTCGCGCCACGCCCCG	1700
624	CCTCGACACGGCCTGGACTAAAGTCACGCTCTGGAAACCGGACATGSGACCTTCGCGCCCA	683
1701	TGGAAACCTCTACGGGTCTCACCCCTTCTACCTTGGCGCTGGAGGACGGCGGTGGGACACA	1760
684	AGTCAACCTCTATGGCTCCCAACCCCTCTACCTTGGTCTAGAGGATGGTGGCTCTGCACA	743
1761	CGGGTGTTCTCTGCTAAACAGCAATGCATGSGATGTGGTCTCTGCAGCCGAGCCCTGCCT	1820
744	TGGTGTCTTCTCTGTAACAGCAATGCCATGSGATGTGCTCTGAGAGCCCTGCCACGCCCT	803
1821	TAGCTGAGGTTCGACAGGTGGGATCTCTGGATGTCTACATCTCTCTGGGCCACAGAGCCAA	1880
804	GACCTGCGCACAAACGGGAGSACTCTCGACTCTTACATCTCTCTGGGCCCTGACCCCA	863
1881	GAGCGTGGTGACAGATPACTCTGAGCTGTGTGGATACCCGTTTCATGCCGCCATCTTGGGG	1940
864	GAGCGTGGTGACAGTACCTGGAATGTGGTGTGGGTTCCCTTCTACGCCCATCTAGGCG	923
1941	CCTGGGCTTCCACTGTGCGCGCTGGGGCTACTCTCCACCGCTATCAACCCGCGAGGTGT	2000
924	CCTGGGCTTCCACTGTGCGCGCTGGGGTTATCTCTCACCTCCACACCCGGCAGGTGC	983
2001	GGAGAACATGACACAGGGCCCCACTTCCCCCTGGAGCTCCAAATGGAAACACCTGAGCTACAT	2060
984	AGCCAACTGTACGTGGGCTCTTCCCTCTGGACGTGCAGTGGAAATGACTGGACTACAT	1043
2061	GGAATCCCGAGGGAATTCAAGTTTCAACAAAGATGGCTTCGGGAACTTCCCGGCCATGGT	2120
1044	GGATGCCAAGAGGGAATTTCACTACAAACAAAGGAAAACCTTCAGAGATTTACCTGCACATGGT	1103
2121	GCAGGAGCTGCACAGGCGCGCGCTACATGATGATCGTGGATCTTCCCATCAGCAG	2180
1104	GCAAGATTTCCATCAGCGTGGCCCTGCATTAAGTCATGATTTGGATCTCTGGCATCAGCAG	1163
2181	CTCGGGCCCTGCGGGAGCTACAGGCCCTACGACGAGGGCTCTCGGAGGGGGGTTTTCAT	2240
1164	CTCGGGCGCGCGGACCTACCGGCCCTATGATGATGGACTGAAGCGAGAGTGTTCAT	1223
2241	CACCAACGACGCGCCAGCGCTGATTTGGAAAGGATATGGCCCGGGTCCACTGCCTTCCC	2300
1224	CCGCAATGCCACGSGGCGAGCGCTCATCGGGAAGGTATGGCCAGGCCCCACGGCTTCCC	1283
2301	CGATTTCAACCAACCCACAGCCCTGGCTCGTGGGAGGACATGGTGGCTGAGTTCCATGA	2360
1284	AGACTTCACCAACCCCTGAGACACACGAGTGGTGGCATGACATGGTGAAGGACTTCACAGA	1343
2361	CCAGTGCCCTTCGACGGCTTTGGATTTGACATGAACGAGCCCTTCAACTTCATCAGAGG	2420
1344	GCAGGTGCCCTTTGACGGGATGTGGATTTGACATGACGAACCATCCNACTTTGTGGAGGG	1403
2421	CTCTGAGGACGGGTGCCCAACAAATGAGCTGGAGAACCCACCTTCAGTGGCTGGGGTGGT	2480
1404	CTCCAGGATGGCTGCCACAGACAGCAGCCCTGGAGAAAGCCCCCGTATGTGCCAGGTGTGT	1463
2481	TGGGGGAGCCTCCAGGCGGCCACCATCTGTGCTCCAGGCCACCGAGTTTCTCTCCACACA	2540
1464	TGGGGGCGGTCTGACGAGCAGGCACCATCTGTGCTCCAGCAGCAGTACCTTTCTCCCA	1523
2541	CTACAACTGCACAACTCTACGGCTGACCGAAGCCATCGCCTCCCAACAGGCGGTGGT	2600
1524	TTACAACTGCACAGCTGTATGGGCTAACTGAGGCCATTTGGCTTCCCACAATGCATGCT	1583
2601	GAAGGCTCGGGGACACGCCCAATTTGTGATCTCCCGCTCGAACCTTTGCTGGCCACGGCCG	2660

1584	DB	CAGGGTGGCGGGCA	CGCGGCCCTT	TGTCATCTCA	CGCTCCAGCTTTGCGGGACACGGGCCA	1644
2661	QY	ATACGCCGGCCACT	TGGACGGGGACG	TGTGGAGCTCCT	TGGAGCAGCTCGCCCTCCTCCGT	2720
1644	DB	TTATGCAGGGCA	CTGGACGGGGAT	TGGAGAGCAGCT	TGGAGCAGCTGSCCCACTCGGT	1703
2721	QY	GCACAGAAATCCT	TGCAGFTTAACT	CTGCTTGGGGGTGCT	TCTGCTCGGGGCCAGCTGTGCGG	2780
1704	DB	GCCAGAGGTGCT	GTCTATTCAACCTCCT	TGTGGGGTTCTCT	TCTGTTGGGAGCTGACATCTGTGG	1763
2781	QY	CTTCTCTGGGC	CAACCTTCAGAGGAGCT	GTGTGTCGCTGACACCGCTG	AGCTGAGCTGGGGCCCTTCTA	2840
1764	DB	CTTTGCTGGTG	CACATCTCAGAGAGCT	TGTGCGCTGACCCAGCT	TGGGACCTTCTA	1823
2841	QY	CCCCTTCATCGG	AAACCAACACGCTCT	CAGTCTGCTCCCGACGAGCG	GTACAGCTTCAG	2900
1824	DB	CCCCTTCATGAG	GAACCAACGACGCGGA	ACCGGCCACAGAGGCCCT	TACGCCCTTCAG	1883
2901	QY	CGAGCGGCCCAG	GAGGCATGAGGAAGG	CCCTCA	CCCTGCGCTACGCACTCTCTCCCCCA	2960
1884	DB	TCTGCTTGCA	CAGGATGCCATGAGG	ATGCTCCCTCCGCTACT	CCTCCTCTCGGCCCTA	1943
2961	QY	CCCTACACACT	GTTCACCAAGCCCA	CGTCGCGGGGAGACG	TGGCCCGGCCCTCTT	3020
1944	DB	CCTTACACCTCT	TTCACCGGGGACACAT	TGGCTGGGACACGGT	TGGCACCGGCCACTGTT	2003
3021	QY	CCTGGAGTCTCC	CAAGGACTCTAGACCT	TGGACTGGACACAGCT	CTCTGTGGGGGGA	3080
2004	DB	CCTTGAGTTCTCC	CAAGACCCCAACACAT	TGGATTTGTGGACCGC	AGCTGTGTGGGGTGC	2063
3081	QY	GGCCCTGCTCAT	CACCCAGTGTCTCAGG	CCCGGGAAGGCCGAAGT	AGTAGCTGGCTACTTTCCC	3140
2064	DB	AGGGCTGCTGAT	CACACCACTGTCTGGA	ACAGGGACAGACCAAA	AGTCAGTGGCTATTTTCCC	2123
3141	QY	CTTGGGCACAT	GTGTACGACTGTGCAG	ACGGTGCCTAATAGAGCCCT	TTGGCAGCCTCCCA	3200
2124	DB	AGTGGGACAT	GGTACAGCTT-----	-----	-----	2144
3201	QY	CCCACCTGCAG	CTCCCCGTGAGCAGC	ATCCACAGCGAGGGG	CAGTGGGTGAGCGTGCC	3260
2145	DB	-----	-----	-----	-----	2192
3261	QY	GGCCCCCTTGG	ACACCATCAACGT	TCACCTCGGGGTGGGTA	CATCATCCCCCTTGCAAGG	3320
2193	DB	AGCCCCCTTGG	ACACCATTAATGT	GCACATCCGAGCAGGG	CACATCTCA	2252
3321	QY	CCCTGGCTCAT	CAACCAACAGTCCG	CCAGCAGCCCATG	CGCTGTGGCCCTAAC	3380
2253	DB	GCCTGGATTGA	ACA	CGGCTGAGTCCCGCA	AGAGGGGATGATGGTGGTGTGGCCCTGAC	2312
3381	QY	CAAGGGTGGAG	AGGCCCGAGCGGAGCT	GTCTTGGGACGATGGAGAG	AGCCCTGGAAGTGCT	3440
2313	DB	ACGGATGGCTT	TGGCGGTGGAGAGCT	GTCTCGGATGATGGG	GAGAGCTGGCAGAGCTT	2372
3441	QY	GGAGCGAGGG	CCCTACACACAGT	CTATCTTCTTGGCCAGGA	ATAACAGATCGTGAA	3500
2373	DB	TGAAGAGGGG	CACTGCAC	TGAGATCTCTCTGGGT	GCACGCTGGTCCGTCTAGTCA	2432
3501	QY	GCTGGTACGT	TGACCAAGTGAAGG	AGCTGGCTGCAGCT	GCAGAGGTGACTCTCC	3560
2433	DB	GATCTTCGGG	CAAGTGGCCACTCTGA	TGGATTCCTGGTGGAGG	CTGTCTGTGTGGTGGG	2492
3561	QY	CGTGGCCAC	CGCGCCCGCAGCAGCT	CTCTCCAA	CGGTGTCCCTGTCTCCAACTTCACCTA	3620
2493	DB	TGTCCCAAGT	CTCCACAGCGAGT	CTCTGGCCAT	TGGCATCCCGGTGGAGAGCTTCTCATA	2552
3621	QY	CAGCCCCG	CAACCAAGTCTTGACAT	CTGTGTCTCGCTGTGT	GATGGGAGAGCAGTTTCT	3680
2553	DB	CCGCAGTGA	CACCCAGGTGCTCGT	GTCTCCGTGTCTG	CTGCCCATGTGGGAGCAGTTTGT	2612
3681	QY	CGTCAGCT	GGTGTTGA	3696		
2613	DB	GGTCCCT	TTGGTCTTGA	2628		

```
RESULT 14
AB000967
LOCUS
DEFINITION
  AB000967 Coturnix japonica GAAI mRNA linear VRT 24-FEB-1999
  cds.
ACCESSION
  AB000967
VERSION
  AB000967.1 GI:3097289
KEYWORDS
  acid alpha glucosidase; GAAI.
SOURCE
  Coturnix japonica (Japanese quail)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
  Phasianinae; Coturnix.
REFERENCE
  1 (sites)
  Kunita,R., Nakabayashi,O., Wu,J.Y., Hagiwara,Y., Mizutani,M.,
  Pennybacker,M., Chen,Y.T. and Kikuchi,T.
  Molecular cloning of acid alpha-glucosidase cDNA of Japanese quail
  (Coturnix coturnix japonica) and the lack of its mRNA in acid
  maltase deficient quails
  Biochim. Biophys. Acta 1362 (2-3), 269-278 (1997)
JOURNAL
  98201735
MEDLINE
  9540858
PUBMED
  2 (bases 1 to 3548)
REFERENCE
  Kikuchi,T.
  Direct Submission
  Submitted (10-FEB-1997) Takeki Kikuchi, National Institute of
  Neuroscence, NCNP, Department of Animal Models for Human Disease;
  4-1-1 Ogawahigashimachi, Kodaira, Tokyo 187, Japan
  (E-mail:tkiku@ncnp.go.jp. Tel:0423-41-2711,
  Fax:0423-46-1754)
FEATURES
  Location/Qualifiers
    1..3548
      /organism="Coturnix japonica"
      /mol_type="mRNA"
      /db_xref="taxon:93934"
      /clone="pQAM8"
      /sex="male"
      /tissue_type="liver"
      /dev_stage="40 day old"
      /note="synonym: Coturnix coturnix japonica"
    1..3548
      /gene="GAAI"
      544..3342
      /gene="GAAI"
      /codon_start=1
      /product="acid alpha glucosidase"
      /protein_id="BAA25884.1"
      /db_xref="GI:3097290"
      /translation="MRSYQKLRATVPQVLVSQEEBEBDAITPTAGRSKLAPWVVG
      SGLITAVLSTITVWLVQVSGWHPAPPAQCHLVPSHRYDCYPERNVVVTQELC
      ERGCCFVITLFAVGKRGVPCFPDPFSPVVSQVSLNQTVMGLLVREKAYPK
      LQMLRMDFQNTKRLHKITIDANPRYEVLEPRVTVKRAENPIYLSLEISQDPFGV
      LRRQGTFTVLLNTVAPLIFADQFLQISTITLPSRLYGLGHEHSTLHSLDWNLTLL
      WARDAPTESFNLVGAHPFLYLLMEGGDAGHVLNLSNAMEVALAPGLTWRTICGV
      LDFYIPLGPDNPMVOQYQEVIFGPAMPPLWALGHLCRWGYGSGNETWOTAKAMRNF
      QIPQADWINDYMGYRDFTPQPKFASLPSLVEDLKHGOHYVILIDPDGISSTSPR
      GYWPFDEGLRGLFNTYQQTLLIQVWPGTATPDFNSNTTHQWLENLQRPHTHV
      PFDGLWIDNEPSNFDSEEGCPGELDSPPTFAVLGNSLTAKTVCASAEQNASVH
      YNLHNLGLKAEATASALIRGRKPFVIRISSTPFSQGRYSGHMLGDNRSQMKDYY
      STPGMLSFSLGFIPLVADICGFSGSEELCTRWMLGAFYPPFRNHNQNEKAQDP
      TAFSPARTAMKDALTRYSLPLVTLFPHRAHQETVARPLFFEPMDVATYGLDR
      QELWGSLLVTPVLEPGADSVLGYRPGQVWYDFYTGSSVSSGEMLKLSAPLDHNLH
      LREGSLPTQKQGIKATRGAPLHLI VALSTRATANGDLFWDDGESLDTFEOGNYSY
      LVFNATENIPTSNVLHASTATDVTIDAVSFYGVQEPSPKVLVDGQEKFPFLYDNLQVL
      TVSGGLVLSQGFSLQWL"
      3548
      /gene="GAAI"
      /note="21 A nucleotides"
  polyA_site
    3548
  ORIGIN
    1121 ACAAGGCCATCACCCAGGACAGTGGGAGCGCGCGCTGTCTACATCCCTGCAAGGC 1180
    DB      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    824 ACGTGGTGGTACCCAGGAGCTGTGCGAGAGCGGGGTGCTGCTTCATCCAGACCCCTCC 883
    QY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    1181 AGGGGCTGCAGGGAGCCAGATGGGGCAGCCCTGTGTCTTCTCCACCCAGCTACCCCA 1240
    DB      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    884 CTGCACT--GGGGGCAAGCGGGGGTGCCTGTGTCTTACCCACCCGACTTCCCA 940
    QY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    1241 GCTACAAGCTGGAGAACCTGAGCTCTCTGAAATGGGCTACACGGCCACCTTGA 1300
    DB      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    941 GCTACGTGGTCAGAGCTCTAACACAGACAGTGTGGGCATGACGGGGTGTGTGGTGC 1000
    QY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    1301 CCACCCCCACCTTCTTCCCAAGGACATCTGACCTGGGCTGGAGCTGATGATGAGA 1360
    DB      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    1001 GGGAAAAAGCCCTATTACCCCAAGGACATCCAGATCTGAGGATGATGATGGAGTTC 1060
    QY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    1361 CTGAGAACCGCTCCACTTTCAGATCAAGATCAAGATCAAGATCAAGATCAAGATCA 1420
    DB      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    1061 CAAACACCGGCTGCACATCAAGATAA CTGATGACAGCAATCCCGCTATGAGGTGCC 1120
    QY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    1421 TGGAGACCCCGCTGTCCACAGCCGGGACCGTCCCACTCTACAGCGTGGAGTCTCCG 1480
    DB      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    1121 TGGAAAGTTCCCGAGTGCACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1180
    QY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    1481 AGGAGCCCTTGGGGGTGATCGTGCACCGGAGCTGGAGCGCGCTGTCTGTAACAGA 1540
    DB      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    1181 AGGACCCCTTGGGGGTGCTGTGAGCGGCAAGGACAGGAGACAGTGTCTGCTCAAC 1240
    QY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    1541 CGGTGGCGCCCTGTCTTGTGGGACCAAGTTCTTCTCAGCTGTCCACCTGTCTGCC 1600
    DB      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    1241 CGGTGGCTCCCTGATCTTGGCCGATCAGTCTCTGCAGATCTCCACACGCTGCC 1300
    QY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    1601 AGTATATACAGGCTCTGCGCGAGACCTCAGTCTCCCTGATGTCTAGCAGCAGCTG 1660
    DB      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    1301 GGTTCCTGTATGGGCTGGGGGAGACCCGAGCAGCACCTGTGTCAGACGCTCGAC 1360
    QY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    1661 GGATCACCTGTGGAAACCGGGACCTTGGCCCGCCAGCCCGGTGCGAAGCTCTAC 1720
    DB      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    1361 CACTCACCTCTGGGCCCGTGTATGTGGCCCGCCAGCGAGTCTTCACTGTACGG 1420
    QY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    1421 ACCCTTCTACCTGTCTGATGGAGGAGGCGCGGATGCCATGGGTTTTCTCTCT 1480
    DB      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    1781 GCAATGCCATGATGTGTCTGACCGAGCCCTGCCCCCTTAGCTGGAGTGCAGAGTG 1840
    DB      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    1481 GCAATGCCATGAGGTGGCCCTGCGACGCTGCCCGCGGCTTGACCTGGAGAAC 1540
    QY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    1841 GGATCCTGATGTCTACATCTTCTGGGCGCCAGAGCCCAAGAGCGTGGTGACAGT 1900
    DB      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    1541 GTGTCTGATCTTCTACATCTTCTGGGCGCTGACCCCAACATGATCTCAGCAGT 1600
    QY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    1601 AGGAGTGTATAGTTTCCAGCCATGCGGCCCTCTGGGCACTTGGCTTCCACCT 1660
    DB      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    1901 TGGAGCTGTGGGATACCCGTTTCTATGCCCCATACTGGGGCTGGGCTTCCACCT 1960
    QY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    1601 AGGAGTGTATAGTTTCCAGCCATGCGGCCCTCTGGGCACTTGGCTTCCACCT 1660
    DB      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    1961 GCTGGGGCTACTCTCCACCGCTATCACCGCCAGGCTGGTGAGAACATGACCA 2020
    QY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    1661 GCTGGGGTGTACGGCTCCAGCAATGAGACCTGGCAGACTGCAAGAGCCACTTCC 1720
    DB      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    2021 ACTTCCCTCTGAGCTGCAATGGAACGACCTGCACTACATGGACTCCCGAGAGG 2080
    QY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    1721 AGATCCACAGAGATGCGCAATGGAACGACATCGATTATATGACCGGGTACCG 1780
    DB      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    2081 CGTTCAACAGAGATGCTTCCGGGACTTCCGGGCGCATGTGTCAGGAGCTGCA 2140
    QY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    1781 CTTTTGATCCCCAGAAAATTTGGCTCCCTTCACTGTGTGGAAGATCTTCCACA 1840
    DB      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    2141 GCGGGGCTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2200
    QY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    26.9%; Score 994.4; DB 5; Length 3548;
    Query Match
```

```
Db 1841 GCGAGCACTACGTGATAATCTTTGGACCCCGGCGATCAGACGACCGACCCACCGTGGCTCT 1900
Qy 2201 ACAGGCGCTTACGACGAGGGTCTCGGAGGGGGTTCATCACCACGAGAGACGGCGACG 2260
Db 1901 ACTGGCTTTCGATGAGGCTTGAGACGGGTTTGTTCCTAAACACACCCAGGGGCGA 1960
Qy 2261 CGCTGATCGGGAAGGATATGCGCGGGTCACTGCGCTTCCCGACTTCACAAACCCACAG 2320
Db 1961 CGCTGATCGGCGAGGTCGTGGCTGGTTACACTGCTTACCTGCTTACCCGACTTCTCAAACAGACA 2020
Qy 2321 CCTGGCTGCTGGGAGGACATGGTGGCTGAGTTCATGACGAGTGCCTTCGACGGCT 2380
Db 2021 CATCATGCTGGTGGGAGAACCTCGAGGCTTCCACACCCACGCTACCTTCGATGGCC 2080
Qy 2381 TGTGATTCACATGAACGAGCGCTTCCAACTTCATCAGAGGCTCTGAGGACGGCTGCCCA 2440
Db 2081 TCTGGATCGACATGATGAGCCATCCAACTTCATGATGGTCTGAAGAGGCTGCCCGC 2140
Qy 2441 ACAATGAGCTGAGAAACCAACCTTACGTGCTGGGGTGGTTGGGGGAACCTTCCAGGGG 2500
Db 2141 CTGGAGAGCTCGACAGCCCGGTACACCGCGCTGTGCTGGGCAATTCCTGACTGCCA 2200
Qy 2501 CCACCATCTGCTCCAGCCACGATTTCTCTCCACACACTACACCTGCAACCTCT 2560
Db 2201 AGACGGTGTGCTCAGCGGAGCAGACGCAATCGGTGCACTACAACTGCAACCTCT 2260
Qy 2561 ACGGCTGACCGAAGCCATCGCTCCACAGGGCGCTGGTGAAGGCTCGGGGACACGCG 2620
Db 2261 ACGGCTGAGAGGACCGAAGCGACGCGAGCGCTTGATCCGGATCCAGGAGGACGCT 2320
Qy 2621 CATTTGTGATCTCCGCTCGACCTTTGCTGGCCACGCGGANTACGCGGCCACTGGAAG 2680
Db 2321 CCTTTGTCATCTCCGCTCCACCTTCCCGAGCCAGGCTCGGTACTCGGGTCACTGGCTGG 2380
Qy 2681 GGAAGCTGTGAGCTCCTGGAGCAGCTCGCTCTCGTGCAGAAATCTTGCAGTTTA 2740
Db 2381 GTGACAAAGGAGCCAAATGGAAGGACATGATTAATCTCCATCCAGGGATGCTGAGCTTCA 2440
Qy 2741 ACTGCTGGGGGCTCTGCTGGGGCGAGCTGCTGGGCTCTCGGCTTCTGGGCAACACCTCAG 2800
Db 2441 GCTGTTTGGATCCCACTGCTGGTGGGGCGGACATCTGTGGCTTCTTGGCAGCACTCGG 2500
Qy 2801 AGGAGCTGTGTGCTGCGCTGGAACCCAGCTGGGGCTTCTACCCCTTCATGCGGAACCA 2860
Db 2501 AGGAGCTGTGCAACCGCTGATGACGCTCGGGCTTCTACCCCTTCTCCCGCAACCA 2560
Qy 2861 ACGCTGTCTGCTGCTGCGGAGGAGCGTACGCTTTCAGGAGCGGGCCGAGCAGGCA 2920
Db 2561 ACAACAGAAACGAGAAAGGCGGAGGACCCACAGCTTTTCAAGCCCTCGGGCAGGACAGCCA 2620
Qy 2921 TGAGGAAGGCGCTCACCTCGCTAGCGACTCTCCCGGACCTCTACACACTGTTCCACG 2980
Db 2621 TGAGGAGCGACTGCTCACCGGCTACTCCCTCTGCTTCTTCTTCTTACAGCTTTTCCATC 2680
Qy 2981 AGGCCCACTGCTGGGGGAGACCGTGGCGCGGCTCTTCTCTGGAGTTTCCCAAGGACT 3040
Db 2681 GTGCCCACTGCAAGGGAGACCGTGGCGGCTTTGTTCTTGTAGTTTCCCTGGGATG 2740
Qy 3041 CTAGCACTGTGGAACCTGAGCAACAGCTCTGTGGGGGAGGCGCTGCTCATCACCCAG 3100
Db 2741 TGCGCACATACGCGCTGAGCAGGAGTTCCTGTGGGGGAGAGCTGCTGTTGACACCGG 2800
Qy 3101 TGCTCAGCGCGGAGGAGGAGTCACTGCTGCTTCCCTTTGGGCAATGTTAGC 3160
Db 2801 TGTGAGCGCGGAGCAGACTCAGTGTGGGCTATTTTCCCGCAGGGGCTGTGGTACAT 2860
Qy 3161 TGCAAGCGGTGCCAATAGAGGCGCTTGGAGCGCTTCCACACCCCGGAGCTTCCCGTG 3220
Db 2861 TCTACAG-----G 2869
Qy 3221 AGCGACCATTCACAGCGAGGGGAGTGGGTGACGCTGGCGGCGCCCTCGGACCATCA 3280
Db 2870 GCTCCTCTGTGAACAGCAGTGGGGAGATGTTAAAGCTGTAAAGCTGTGAGCGGCCCTCGGACCATCA 2929
```

```
Qy 3281 AGTCCACTCTCGGCTGGGTACATCATCCCTTCAGAGGCGCTTGGCTTCAACACACAG 3340
Db 2930 ACCTGCATCTCGGAGGGTTCATCTCTGCCACACAGAAACCGGGATCACAGCAAGG 2989
Qy 3341 AGTCCCGCAGCAGCCCATGGCTTGGCTTGGCCCTTAACCAAGGTGGAGAGGCCGAG 3400
Db 2990 CCACCCGAGGGAACCTCTGATCTCATTTGGGCCCTGTCCACACGTGCCACTGCTGGG 3049
Qy 3401 GGGAGCTGTTCTGGGACGATGGAGAGAGCTGGAAAGTGTGGAGGAGGGGCTTACACAC 3460
Db 3050 GGGACCTTTCTGGGATGATGGAGAGAGCTGGACACCTTCGAGCAGGGGCAACTACTCT 3109
Qy 3461 AGGTCACTTCTCTGGCCAGGAATAACAGATCGTGAATAGCTGGTACGTGTGACAGTG 3520
Db 3110 ACCTGGTGTCAATGCCAGGAGAACTTTTCCACCTCCAACGCTCTTATGCGACGACCG 3169
Qy 3521 AGGAGCTGGCTGAGCTGCGAGAGGTGACTGTCTGGGCTGGCCAGCGGCCCGCCAGC 3580
Db 3170 AGGCCACCGATGTACCATCGATGAGTGTCTATGAGTGCAGGAGCGGCCCGCAGCA 3229
Qy 3581 AGGTCTCTTCCAAACGG 3596
Db 3230 AGTCTCTCTGATGG 3245
```

RESULT 15

```
LOCUS Q0776406 1493 bp DNA linear PAT 11-MAR-2004
DEFINITION Sequence 92 from Patent EP1394274.
```

```
ACCESSION Q0776406
```

```
VERSION Q0776406.1 GI:45379796
```

```
KEYWORDS Homo sapiens (human)
```

```
SOURCE
```

```
ORGANISM
```

```
Homo sapiens
```

```
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE 1
```

```
AUTHORS Ohtani,N., Sugita,Y., Yamaya,M., Kubo,H., Nagai,H. and Izuohara,K.
```

```
TITLE Methods of testing for bronchial asthma or chronic obstructive
```

```
JOURNAL Patent: EP 1394274-A 92 03-MAR-2004;
```

```
Genex Research, Inc. (JP)
```

```
FEATURES
```

```
source
```

```
1..1493
```

```
/organism="Homo sapiens"
```

```
/mol_type="unassigned DNA"
```

```
/db_xref="taxon:9606"
```

```
ORIGIN
```

```
Query Match 25.9%; Score 957; DB 6; Length 1493;
Best Local Similarity 99.0%; Pred. No. 1.5e-138;
```

```
Matches 963; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
```

```
Qy 63 CGGATCTTACTCGCGGAGAAAGAACAGCCCAAGCGTCCCGGAAACGCGAGTCCGGAGA 122
```

```
Db 109 CGGCAAGTACTCGCGGAGAAAGAACAGCCCAAGCGTCCCGGAAACGCGAGTCCGGAGA 168
```

```
Qy 123 GGAGTTCGCGATGAGAAAGTTGAACCGCTGTGGGAGAGGCCCGGACTGCATCTTCC 182
```

```
Db 169 GGAGTTCGCGATGAGAAAGTTGAACCGCTGTGGGAGAGGCCCGGACTGCATCTTCC 228
```

```
Qy 183 TCCCGTGAAGCTGGCGGAGCTCCAGCTGATCTGAAGATACAGGAGAGGAGCGAACTCCG 242
```

```
Db 229 TCCCGTGAAGCTGGCGGAGCTCCAGCTGATCTGAAGATACAGGAGAGGAGCGAACTCCG 288
```

```
Qy 243 CTGGAAGAACTAAAGCTTGACGGCTTGAACGAGATGGGGAGAAAGAGCGAGACTCAT 302
```

```
Db 289 CTGGAAGAACTAAAGCTTGACGGCTTGAACGAGATGGGGAGAAAGAGCGAGACTCAT 348
```

```
Qy 303 AGCGAACTCAATGTATCTTGGCCAAAGTATGGTCTGGACGGAAGAGAGCGCTCGGCA 362
```

```
Db 349 AGCGAACTCAATGTATCTTGGCCAAAGTATGGTCTGGACGGAAGAGAGCGCTCGGCA 408
```

```
QY 363 GGTGACCAAGCACTCCTCAGTGGCACCCAGGAAGACGGGCTGGATGACCCCAAGGCTGGA 422
Db |||||
QY 409 GGTGACCAAGCACTCCTCAGTGGCACCCAGGAAGACGGGCTGGATGACCCCAAGGCTGGA 468
Db |||||
QY 423 AAAGCTGTGGCACAAGCGGAAGACCTCTGGGAAATTTCTCGGCGAAGAACTGGACAAGCT 482
Db |||||
QY 469 AAAGCTGTGGCACAAGCGGAAGACCTCTGGGAAATTTCTCGGCGAAGAACTGGACAAGCT 528
QY 483 CTGGCGGGAGTTCTCTCATTCACAAAGAGAAAGTTACAGAGTACAAAGTCTCTGCTGGAGAC 542
Db |||||
QY 529 CTGGCGGGAGTTCTCTCATTCACAAAGAGAAAGTTACAGAGTACAAAGTCTCTGCTGGAGAC 588
QY 543 CTTGAGCAGGACCGAAGAAATCCACAGAGAACGTCATTAGCCCTCGGACCTTGAGCGACAT 602
Db |||||
QY 589 CTTGAGCAGGACCGAAGAAATCCACAGAGAACGTCATTAGCCCTCGGACCTTGAGCGACAT 648
QY 603 CAAGGCGAGGCTCTCTCACAGCAGGACACACGAGCTGAAGGAGAACTGGCGAGCATCAA 662
Db |||||
QY 649 CAAGGCGAGGCTCTCTCACAGCAGGACACACGAGCTGAAGGAGAACTGGCGAGCATCAA 708
QY 663 CCAAGGCGCTGGACCCGCTTGGCAGGCTCAGCCACAGGCTACAGCACTGAGGCTGAGTT 722
Db |||||
QY 709 CCAAGGCGCTGGACCCGCTTGGCAGGCTCAGCCACAGGCTACAGCACTGAGGCTGAGTT 768
QY 723 CGAGGAGCCCGAGGCTGATTGACCTGTGGACCTTGGGACCTTGGCGCAGTCCGCCAACTCACGGACAA 782
Db |||||
QY 769 CGAGGAGCCCGAGGCTGATTGACCTGTGGACCTTGGGACCTTGGCGCAGTCCGCCAACTCACGGACAA 828
QY 783 GGAGCTGGAGGCTTCCGGGAGGAGCTCAAGCACTTCGAAGCCAAATCGAGAAGCAAA 842
Db |||||
QY 829 GGAGCTGGAGGCTTCCGGGAGGAGCTCAAGCACTTCGAAGCCAAATCGAGAAGCAAA 888
QY 843 CCACTACCAAGACAGCTGGAGATTGCGCACGAGAAAGCTGAGGCACGACAGAGAGCTGGG 902
Db |||||
QY 889 CCACTACCAAGACAGCTGGAGATTGCGCACGAGAAAGCTGAGGCACGACAGAGAGCTGGG 948
QY 903 CGACGGCGAGCGTGTGAGCGCGCAGCGCGAGAAAGCAAGCTTCTGCTGGAGGGCGGACCAA 962
Db |||||
QY 949 CGACGGCGAGCGTGTGAGCGCGCAGCGCGAGAAAGCAAGCTTCTGCTGGAGGGCGGACCAA 1008
QY 963 GGAGCTGGGCTACACGGTGAAGAGCATCTCGAGGACCTTCCGCGAGGATCTCCAGAGC 1022
Db |||||
QY 1009 GGAGCTGGGCTACACGGTGAAGAGCATCTGAGGACCTTCCGCGAGGATCTCCAGAGC 1068
QY 1023 TCGCGCCCGAGGCA 1035
Db |||
QY 1069 TCGGCACAACGAA 1081
Db |||
```

Search completed: April 4, 2005, 15:52:18
Job time : 15278 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 4, 2005, 11:07:54 ; Search time 1756 Seconds
(without alignments)
12480.001 Million cell updates/sec

Title: US-10-600-862A-6
Perfect score: 3702
Sequence: 1 ctaccgccatcggggtcc.....tcagtggtggtgactcgag 3702

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_16Dec04:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002as:*
- 7: Geneseqn2002bs:*
- 8: Geneseqn2003as:*
- 9: Geneseqn2003bs:*
- 10: Geneseqn2003cs:*
- 11: Geneseqn2003ds:*
- 12: Geneseqn2004as:*
- 13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2647.8	71.5	3026	13 ADQ91897	Adq91897 Human glu
2	2647.8	71.5	3624	10 ACf80584	Acf80584 Human alp
3	2647.8	71.5	3846	13 ADQ91895	Adq91895 Human glu
4	2636.6	71.2	3847	13 ACn40294	Acn40294 Tumour-as
5	2635.4	71.2	2851	12 ADf47520	Adf47520 Human GAA
6	2186.2	59.1	3376	13 ACn42967	Acn42967 Human dia
7	2097.2	56.7	3455	13 ACn42966	Acn42966 Human dia
8	1887.4	51.0	3105	13 ACn42968	Acn42968 Human dia
9	1881.4	50.8	3364	6 ABi99468	Abi99468 Mouse isc
10	1191.6	32.2	2655	13 ACN42969	ACN42969 Human dia
11	957	25.9	1493	5 AAC87527	Aac87527 Human RAP
12	957	25.9	1493	6 AAL48893	Aal48893 Human rec
13	957	25.9	1493	6 ABK84062	Abk84062 Human cdn
14	957	25.9	1493	12 ADJ74840	Adj74840 Marker ge
15	957	25.9	1493	13 ADR24896	Adr24896 Breast ca
16	957	25.9	1493	13 ADP23134	Adp23134 PRO:polyd
17	954.2	25.8	2084	13 ADQ38777	Adq38777 Human SNP
18	954.2	25.8	2106	13 ADQ38776	Adq38776 Human SNP
19	953.8	25.8	1499	2 AA233614	Aaz33614 Human bre
20	665.6	18.0	2902	6 ABi99326	Abi99326 Mouse isc

c	21	665.6	18.0	2902	12	ADJ75755	Adj75755 Marker ge
	22	437.2	11.8	470	9	ACH40400	Ach40400 Human foe
	23	411	11.1	956	6	ABN74475	Abn74475 Bovine em
	24	372.4	10.1	421	6	ABV94943	Abv94943 Human pan
	25	371.2	10.0	5825	9	ACH03918	Ach03918 Human cdn
	26	371.2	10.0	6483	4	AAH57440	Aah57440 Human int
	27	371.2	10.0	6513	12	ADO19559	Ado19559 Human PRO
	28	370.6	10.0	449	6	ABT07013	Abt07013 Human ova
	29	370.6	10.0	449	8	ABX72891	Abx72891 Human ova
	30	369.8	10.0	2752	2	AAV11736	Aav11736 Barley al
	31	368	9.9	1644	10	ADB53939	Adb53939 Primary r
	32	368	9.9	1644	10	ABT42500	Abt42500 Toxicity
	33	368	9.9	1644	12	ADP73017	Adp73017 Renal tox
	34	361	9.8	991	6	ABN74476	Abn74476 Bovine em
	35	338.8	9.2	18272	4	AAK69447	Aak69447 Human imm
	36	338.8	9.2	18272	4	AAJ06181	Aaj06181 Human rep
	37	338.8	9.2	18272	4	ABL98746	AbL98746 Human tea
	38	318.6	8.6	466	9	ACH48464	Ach48464 Human leu
	39	298.2	8.1	465	9	ACH18185	Ach18185 Human adu
	40	284	7.7	649	13	ADQ52009	Adq52009 Novel can
c	41	284	7.7	667	10	ADB57541	Adb57541 Toxicity-
	42	284	7.7	667	10	ADB52060	Adb52060 Primary r
c	43	284	7.7	667	12	ADP72364	Adp72364 Renal tox
	44	279.2	7.5	1560	10	ADC07849	Adc07849 Rice DNA
	45	278.6	7.5	1560	10	ADC08230	Adc08230 Rice DNA

ALIGNMENTS

RESULT 1
ADQ91897
ID ADQ91897 standard; cdna; 3026 BP.

XX ADQ91897;

XX 21-OCT-2004 (first entry)

XX Human glucosidase alpha acid (GAA) encoding cdna SEQ ID NO:3.

XX chimeric protein; secretory signal; lysosomal protein;

KW lysosomal acid alpha-glucosidase; glucosidase alpha acid; GAA; enzyme;

KW lysosomal acid alpha-glucosidase deficiency; GAA deficiency;

KW glycogen storage disease type II; GSD II; human; chromosome 17; gene; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS 33..2891

FT /*tag= a

FT /product= "glucosidase alpha acid (GAA)"

XX WO2004064750-A2.

XX 05-AUG-2004.

XX 21-JAN-2004; 2004WO-US001453.

XX 22-JAN-2003; 2003US-0441789P.

XX (UYDU-) UNIV DUKE.

XX Koeberl DD, Sun B;

XX WPI; 2004-571599/55.

DR P-PSDB; ADQ91896.

DR GENBANK; NM_000152.

XX New nucleic acid expressing lysosomal acid alpha-glucosidase (GAA)

PT polypeptide, useful for preparing a composition for treating e.g.,

PT glycogen storage disease type II.

XX Claim 45; SEQ ID NO 3; 128pp; English.


```
Qy 2656 GCGCGATACCGCGCCACTCGACGGGGGAGCTGTGGAGCTCCTGGGAGCAGCTCGCCTCC 2715
Db 1851 GCGCGATACCGCGCCACTCGACGGGGGAGCTGTGGAGCTCCTGGGAGCAGCTCGCCTCC 1910
Qy 2716 TCCGTGCGCAGAAATCCCTGCAAGTTTAACTCTGCTGGGGGTGCTCTGTGGGGGCGGACGTC 2775
Db 1911 TCCGTGCGCAGAAATCCCTGCAAGTTTAACTCTGCTGGGGGTGCTCTGTGGGGGCGGACGTC 1970
Qy 2776 TGGGCTTCCTGGGCAACACCTCAGAGGAGCTGTGTGGCTGGACCCAGCTGGGGGCG 2835
Db 1971 TGGGCTTCCTGGGCAACACCTCAGAGGAGCTGTGTGGCTGGACCCAGCTGGGGGCG 2030
Qy 2836 TTCTACCCCTTTCATGGGACCAACACAGCTCTCAGTCTGCCCCAGGAGCGGTACAGC 2895
Db 2031 TTCTACCCCTTTCATGGGACCAACACAGCTCTCAGTCTGCCCCAGGAGCGGTACAGC 2090
Qy 2896 TTCAGGAGCGCGGCCAGCAGCGCCATGAGGAAGCCCTCACCTCGGCTACGCACTCCTC 2955
Db 2091 TTCAGGAGCGCGGCCAGCAGCGCCATGAGGAAGCCCTCACCTCGGCTACGCACTCCTC 2150
Qy 2956 CCCCACTTACACATCTGTTTCCACAGGCCACGTCGCGGGGAGACCGTGGGCCCGGCC 3015
Db 2151 CCCCACTTACACATCTGTTTCCACAGGCCACGTCGCGGGGAGACCGTGGGCCCGGCC 2210
Qy 3016 CTCTTCTGGAGTTCCCAAGGACTTAGCACCTGAGCTGTGACACACAGCTCTCTGTGG 3075
Db 2211 CTCTTCTGGAGTTCCCAAGGACTTAGCACCTGAGCTGTGACACACAGCTCTCTGTGG 2270
Qy 3076 GGGAGGCGCTGCTCATCACCCAGTGTCTCAGGCGGGAAGCCGAGTGAAGTGAAGTGA 3135
Db 2271 GGGAGGCGCTGCTCATCACCCAGTGTCTCAGGCGGGAAGCCGAGTGAAGTGAAGTGA 2330
Qy 3136 TTCCCTTGGGCAATGTTGACGCTGCGAGACGGTGCCTCAATAGAGCCCTTGGACGCTC 3195
Db 2331 TTCCCTTGGGCAATGTTGACGCTGCGAGACGGTGCCTCAATAGAGCCCTTGGACGCTC 2390
Qy 3196 CGACCCCGCTGAGCTCCCTGAGCGAGCCAGCCATCGACGAGGGGAGTGGGTGAGC 3255
Db 2391 CGACCCCGCTGAGCTCCCTGAGCGAGCCAGCCATCGACGAGGGGAGTGGGTGAGC 2450
Qy 3256 CTGCGGCGCCCTGAGACCAATCAAGCTCACCTCGGCGTGGGTACATATCCCCCTG 3315
Db 2451 CTGCGGCGCCCTGAGACCAATCAAGCTCACCTCGGCGTGGGTACATATCCCCCTG 2510
Qy 3316 CAGGCGCTTGGCTCAACACACAGAGTCCCGCAGCAGCCCATGCGCTTGGCTGTGGCC 3375
Db 2511 CAGGCGCTTGGCTCAACACACAGAGTCCCGCAGCAGCCCATGCGCTTGGCTGTGGCC 2570
Qy 3376 CTAAACAGGCTGAGAGCGCGGAGCTGTCTTGGGACGATGGAGAGCCTGGAA 3435
Db 2571 CTGACCAAGGCTGAGAGCGCGGAGCTGTCTTGGGACGATGGAGAGCCTGGAA 2630
Qy 3436 GTGCTGGAGCGAGGGGCTTACACAGAGTCACTCTCTGCGCAGGAATTAACAGATCGG 3495
Db 2631 GTGCTGGAGCGAGGGGCTTACACAGAGTCACTCTCTGCGCAGGAATTAACAGATCGG 2690
Qy 3496 AATGAGCTGGTACGTGTGACCAAGTGAAGGAGCTGGCTCGAGCTGCAGAGGTGACTGTC 3555
Db 2691 AATGAGCTGGTACGTGTGACCAAGTGAAGGAGCTGGCTCGAGCTGCAGAGGTGACTGTC 2750
Qy 3556 CTGGGCGTGGCCACGGGCGCCAGAGGCTCTCTCAAGGCTGTCTCTTCCAACTTC 3615
Db 2751 CTGGGCGTGGCCACGGGCGCCAGAGGCTCTCTCAAGGCTGTCTCTTCCAACTTC 2810
Qy 3616 ACTCAGAGCCCGACACCAAGGCTCTGGACATCTGTCTCGCTGTGTGATGGAGAGCAG 3675
Db 2811 ACTCAGAGCCCGACACCAAGGCTCTGGACATCTGTCTCGCTGTGTGATGGAGAGCAG 2870
Qy 3676 TTCTCTGTGAGCTGGTGT 3694
Db 2871 TTCTCTGTGAGCTGGTGT 2889
```

RESULT 2

ACF80584
ID ACF80584 standard; cDNA; 3624 BP.

XX ACF80584;

XX AC ACF80584;

XX DT 15-JAN-2004 (first entry)

XX Human alpha-glucosidase coding sequence.

XX DE Human; alpha-glucosidase; lysosome; enzyme; Pompe disease;

XX KW transgenic plant; gene; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX CDS 220..3078

XX FT /tag= a

XX FT /product= "Human alpha-glucosidase"

XX FT sig_peptide 220..426

XX FT mat_peptide 427..3075

XX FT /tag= c

XX WO2003073839-A2.

XX PN 12-SEP-2003.

XX PD 03-MAR-2003; 2003WO-IT000120.

XX PF 01-MAR-2002; 2002IT-RM000115.

XX PR (PLAN-) PLANTECHNO SRL.

XX PA Fogher C, Reggi S;

XX PI WPI; 2003-712829/67.

XX DR P-PSDB; ABM79002.

XX PT New genetically transformed plant that can produce a lysosomal enzyme of

XX PT animal or human origin, useful for preparing a medicament for enzyme

XX PT replacement therapy in Gaucher, Anderson-Fabry or Pompe disease.

XX PS Example 10; Page 49-51; 53pp; English.

XX CC The present sequence is the coding sequence for human alpha-glucosidase

XX CC (GAA). A deficiency of this enzyme causes Pompe disease. The invention is

XX CC based on the discovery that lysosomal enzymes, such as GAA, can be

XX CC expressed in seed storage organs in a form which is stable (over 12

XX CC months in stored seeds), enzymatically active and in a high amount

XX CC PCR using primers that deleted the signal sequence and native polyA site.

XX CC It was used to construct vector pL4200, which also included the basic

XX CC 7S soy globulin promoter (PGLOB, see ACF80579) and the basic 7S soy

XX CC globulin signal sequence (see ACF80580) upstream of the vector's polyA

XX CC site. Such constructs can be used to express human GAA in transgenic

XX CC plants, especially in plants having a high protein content, e.g. legumes,

XX CC cereals and tobacco

XX SQ Sequence 3624 BP; 643 A; 1220 C; 1100 G; 661 T; 0 U; 0 Other;

Query Match 71.5%; Score 2647.8; DB 10; Length 3624;

Best Local Similarity 99.7%; Pred. No. 0;

Matches 2652; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1036 GAAACCGGTGACACCCCGCGGCTCCAGAGAGTGGCCACACAGTGGAGCTCCCCCCC 1095

Db 418 GATGCCACGGACACACCCCGCGGCTCCAGAGAGTGGCCACACAGTGGAGCTCCCCCCC 477

Qy 1096 AACAGCCGCTTCGATTGGCCCTTGACAAAGCCATCACCCAGGACAGTGGAGGCCGC 1155

Db 478 AACAGCCGCTTCGATTGGCCCTTGACAAAGCCATCACCCAGGACAGTGGAGGCCGC 537

QY	1156	GGCTGCTGCTACATCCCTGCAAGCAGGGGCTGACAGAGCCAGATGGGCGAGCCCTGG	1215		Db	1618	TTTCATCAACAACGACAGCGGCGAGCGCTGATTGGGAAGGTATGGCCCGGGTCCATGGC	1677	
Db	538	GGCTGCTGCTACATCCCTGCAAGCAGGGGCTGACAGAGCCAGATGGGCGAGCCCTGG	597		QY	2296	TTTCCCGGACTTACCAACCCCAACAGCCCTGGCTGGTGGGAGGACATGGTGGCTGAGTTTC	2355	
QY	1216	TGCTTTCTTCCACCCAGCTTACCCAGCTACAGCTGAGAACTGAGCTCTCTGAAATG	1275		Db	1678	TTTCCCGGACTTACCAACCCCAACAGCCCTGGCTGGTGGGAGGACATGGTGGCTGAGTTTC	1737	
Db	598	TGCTTTCTTCCACCCAGCTTACCCAGCTACAGCTGAGAACTGAGCTCTCTGAAATG	657		QY	2356	CATGACCAAGGTGCCCTTCGACGGCTTGTGGATTGACATGAACGAGCCTTCCAATTCATC	2415	
QY	1276	GGCTACACGGCCACCTGACCGGTACCAACCCCAACCTTCTCCCAAGGACATCTCTGACC	1335		Db	1738	CATGACCAAGGTGCCCTTCGACGGCATGTGGATTGACATGAACGAGCCTTCCAATTCATC	1797	
Db	658	GGCTACACGGCCACCTGACCGGTACCAACCCCAACCTTCTTCCCAAGGACATCTCTGACC	717		QY	2416	AGAGGCTCTGAGGACGGCTGCCCCAAATGATGAGCTGGAGAACCAACCTTA COTGGCTGGG	2475	
QY	1336	CTGCGGCTGAGAGTGATGAGAGCTGAGAACCGGCTCCACTTCAAGTCAAAAGATCCA	1395		Db	1798	AGAGGCTCTGAGGACGGCTGCCCCAAATGATGAGCTGGAGAACCAACCTTA COTGGCTGGG	1857	
Db	718	CTGCGGCTGAGAGTGATGAGAGCTGAGAACCGGCTCCACTTCAAGTCAAAAGATCCA	777		QY	2476	GTGGTTGGGGGAGCCTCCAGSGGSCCACTCTGTGCTCCAGCAACAAGTTTCTCTCC	2535	
QY	1396	GCTAAACAGGGCTACAGAGTGCCCTTGGAGACCCCGCGGTGTCAACAGCGGGACCGTCC	1455		Db	1858	GTGGTTGGGGGAGCCTCCAGSGGSCCACTCTGTGCTCCAGCAACAAGTTTCTCTCC	1917	
Db	778	GCTAAACAGGGCTACAGAGTGCCCTTGGAGACCCCGCGGTGTCCACAGCGGGACCGTCC	837		QY	2536	ACACACTACAACTTCACAAACCTTACCGGCTTCAGCCGAGCCATCGCCTCCCAAGGGCG	2595	
QY	1456	CCACTCTACAGCGTGGAGTTCTCCGAGGAGCCCTTCGGGCTGATCGTGCAACGGGACGCTG	1515		Db	1918	ACACACTACAACTTCACAAACCTTACCGGCTTCAGCCGAGCCATCGCCTCCCAAGGGCG	1977	
Db	838	CCACTCTACAGCGTGGAGTTCTCCGAGGAGCCCTTCGGGCTGATCGTGCAACGGGACGCTG	897		QY	2596	CTGGTGAAGGCTCGGGGAGACAGCCCATTTGTGATCTCCCGCTCGACCTTTGCTGGCCAC	2655	
QY	1516	GACGGCGCGTGTCTGTAACACAGACGCTGGGCGCCCTGTCTTTTGGGACCAAGTTCTTT	1575		Db	1978	CTGGTGAAGGCTCGGGGAGACAGCCCATTTGTGATCTCCCGCTCGACCTTTGCTGGCCAC	2037	
Db	898	GACGGCGCGTGTCTGTAACACAGACGCTGGGCGCCCTGTCTTTTGGGACCAAGTTCTTT	957		QY	2656	GGCCGATACCGCGGCACTGGACGGGGAGCGTGTGGAGCTCTCTGGGAGCAGCTCCCTCC	2715	
QY	1576	CAGCTGTCCACTCGCTGCGCTCGAGTATATCAACAGGCTTCGCGAGCACTCAGTCCC	1635		Db	2038	GGCCGATACCGCGGCACTGGACGGGGAGCGTGTGGAGCTCTCTGGGAGCAGCTCCCTCC	2097	
Db	958	CAGCTGTCCACTCGCTGCGCTCGAGTATATCAACAGGCTTCGCGAGCACTCAGTCCC	1017		QY	2716	TCGGTCCAGAAATCCTGCAGTTTAACTGTCTGGGGGTGCCTCTGGTGGGGCCGACGTC	2775	
QY	1636	CTGATGCTCAGCACAGCTGGACCAAGATCACCTCTGGAACCGGACCTTCGCCCCAGG	1695		Db	2098	TCCGTGCCAGAAATCCTGCAGTTTAACTGTCTGGGGGTGCCTCTGCTGGGGCCGACGTC	2157	
Db	1018	CTGATGCTCAGCACAGCTGGACCAAGATCACCTGTGGAACCGGACCTTTGGCCCAAG	1077		QY	2776	TGCGGCTTCTGGGCAACAACCTCAGAGGAGCTGTGTGCGTGTGACCAACAGCTGGGGGCC	2835	
QY	1696	CCGGTGCAGAACCTTACCGGCTCACCCCTTTCTACTCGCGCTGGAGGACGGCGGGTGG	1755		Db	2158	TGCGGCTTCTGGGCAACAACCTCAGAGGAGCTGTGTGCGCTGGACCAACAGCTGGGGGCC	2217	
Db	1078	CCGGTGCAGAACCTTACCGGCTCACCCCTTTCTACTCGGCGTGGAGACGGCGGGTGG	1137		QY	2836	TTTCTACCTTTCATCGGGAAACCAACAACAGCTGTCTGCTGSCCCAGAGCCGTACACG	2895	
QY	1756	GCACAGGGGTGTCTCTGCTAAACAGCAATGCCATGGATGTGGTCTGACGGAGCCCT	1815		Db	2218	TTTCTACCTTTCATCGGGAAACCAACAACAGCTGTCTGCTGCCCCAGAGCCGTACAGC	2277	
Db	1138	GCACAGGGGTGTCTCTGCTAAACAGCAATGCCATGGATGTGGTCTCTGACGGAGCCCT	1197		QY	2896	TTTCTACCTTTCATCGGGAAACCAACAACAGCTGTCTGCTGCCCCAGAGCCGTACAGC	2955	
QY	1816	GCCTTTAGCTGGAGGTGCAACAGGTGGATTCCTGGATGTCTACATCTCTGGGCCACAG	1875		Db	2278	TTTCTACCTTTCATCGGGAAACCAACAACAGCTGTCTGCTGCCCCAGAGCCGTACAGC	2337	
Db	1198	GCCTTTAGCTGGAGGTGCAACAGGTGGATTCCTGGATGTCTACATCTCTGGGCCACAG	1257		QY	2956	CCCAACCTTACACACTGTTTCCACAGGCCACAGTTCGGGGGAGACCGTGGCCCGGGCC	3015	
QY	1876	CCCAAGAGCGTGTGACAGGTAACCTGGACGTTGTGGGATAACCGTTTCATGCGGCCATAC	1935		Db	2338	CCCAACCTTACACACTGTTTCCACAGGCCACAGTTCGGGGGAGACCGTGGCCCGGGCC	2397	
Db	1258	CCCAAGAGCGTGTGACAGGTAACCTGGACGTTGTGGGATAACCGTTTCATGCGGCCATAC	1317		QY	3016	CTCTTCTTGGAGTTTCCCAAGGACTTAGCACTGGACTGTGTGGAACAACAGCTCTGTGG	3075	
QY	1936	TGGGGCTTGGGCTTCACTGTGCGCTGGGGTACTCTCTCAACCGCTATACCCCGCCAG	1995		Db	2398	CTCTTCTTGGAGTTTCCCAAGGACTTAGCACTGGACTGTGTGGAACAACAGCTCTGTGG	2457	
Db	1318	TGGGGCTTGGGCTTCACTGTGCGCTGGGGTACTCTCTCAACCGCTATACCCCGCCAG	1377		QY	3076	GGGGAGGCGCTGCTCATCAACCCAGTGTCTCAGGCCGGGAAAGGCCGAAAGTGTAGCTAC	3135	
QY	1996	GTGGTGGAGAACATGACAGGGGCCACTTCCCGCTGGAAGTCCAAATGGAAACGACCTGGAC	2055		Db	2458	GGGGAGGCGCTGCTCATCAACCCAGTGTCTCAGGCCGGGAAAGGCCGAAAGTGTAGCTAC	2517	
Db	1378	GTGGTGGAGAACATGACAGGGGCCACTTCCCGCTGGAAGTCCAAATGGAAACGACCTGGAC	1437		QY	3136	TTTCCCTTGGGCAACATGGTACGACCTGACAGCGGTGCCAATAGAGGCCCTTGGCAGCCTC	3195	
QY	2056	TACATGGACTCCCGAGGAGCTTCAAGTTCAACAGGATGGCTTCGGGACTTCCCGGCC	2115		Db	2518	TTTCCCTTGGGCAACATGGTACGACCTGACAGCGGTGCCAATAGAGGCCCTTGGCAGCCTC	2577	
Db	1438	TACATGGACTCCCGAGGAGCTTCAAGTTCAACAGGATGGCTTCGGGACTTCCCGGCC	1497		QY	3196	CCACCCCAACCTGACGCTCCCGTACAGCGGTGCCAATAGAGGCCCTTGGCAGCCTC	3255	
QY	2116	ATGGTGCAGAGCTGACACAGGGCGCGCGCTACATGATGATCTGTGATCTCTGCCATC	2175		Db	2578	CCACCCCAACCTGACGCTCCCGTACAGCGGTGCCAATAGAGGCCCTTGGCAGCCTC	2637	
Db	1498	ATGGTGCAGAGCTGACACAGGGCGCGCGCTACATGATGATCTGTGATCTCTGCCATC	1557		QY	3256	CTGCGGGCGCCCTCGGACACCATCAACGCTTCAACCTCCGGGCTGGGTATATCAATCCCCCTG	3315	
QY	2176	AGCAGCTCGGGCCCTCGCGGAGCTACAGGCCCTTACAGAGGGTCTCGGAGGGGGGT	2235		Db	2638	CTGCGGGCGCCCTCGGACACCATCAACGCTTCAACCTCCGGGCTGGGTATATCAATCCCCCTG	2697	
Db	1558	AGCAGCTCGGGCCCTCGCGGAGCTACAGGCCCTTACAGAGGGTCTCGGAGGGGGGT	1617		QY	3316	CAGGGCGCTGGGCTCAACAACAGAGTCCCGCCAGAGCCCATGGCCCTTGGCTGTGGCC	3375	
QY	2236	TTTCATCAACAACAGAGACCGGCCAGCCGCTGATTTGGGAAGGTATGGGCCCGGGTCCACTGCC	2295						

Db 2698 CAGGGCCCTGGCTCACAACACAGAGTCCCGCAGAGCCCATGGCTGTGGCC 2757
Qy 3376 CTAACAAGGGTGGAGAGGCCCGAGGGAGCTGTCTGGACGATGGAGAGCCTGGAA 3435
Db 2758 CTGACCAAGGGTGGAGAGGCCCGAGGGAGCTGTCTGGACGATGGAGAGCCTGGAA 2817
Qy 3436 GTGCTGGAGGGGGCTTACACAGGTATCTTCTGGCCAGGAAATACAGATCGTG 3495
Db 2818 GTGCTGGAGGGGGCTTACACAGGTATCTTCTGGCCAGGAAATACAGATCGTG 2877
Qy 3496 AATGAGCTGTAGCTGTGACCTGAGGAGGCTGCGCTGAGCTGCAGAGGTGACTGTC 3555
Db 2878 AATGAGCTGTAGCTGTGACCTGAGGAGGCTGCGCTGAGCTGCAGAGGTGACTGTC 2937
Qy 3556 CTGGGGCTGGCCACGGCGGCCAGCAGGCTCTCTCAACGGTGTCTCTCAACTTC 3615
Db 2938 CTGGGGCTGGCCACGGCGGCCAGCAGGCTCTCTCAACGGTGTCTCTCAACTTC 2997
Qy 3616 ACCTACAGCCCCGACACCAAGTCTCTGGACATCTGTCTGCTGTGATGGAGAGCAG 3675
Db 2998 ACCTACAGCCCCGACACCAAGTCTCTGGACATCTGTCTGCTGTGATGGAGAGCAG 3057
Qy 3676 TTTCTCGTCAGCTGGTGT 3694
Db 3058 TTTCTCGTCAGCTGGTGT 3076

RESULT 3

AD091895
ID AD091895 standard; cDNA; 3846 BP.
XX AC AD091895;
XX DT 21-OCT-2004 (first entry)
XX DE Human glucosidase alpha acid (GAA) encoding cDNA SEQ ID NO:1.
XX KW chimeric protein; secretory signal; lysosomal protein;
KW lysosomal acid alpha-glucosidase; glucosidase alpha acid; GAA; enzyme;
KW lysosomal acid alpha-glucosidase deficiency; GAA deficiency;
KW glycogen storage disease type II; GSD II; human; chromosome 17; gene; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT CDS 442..3300
FT FT /*tag= a
FT FT /product= "glucosidase alpha acid (GAA)"
XX WO2004064750-A2.
XX PD 05-AUG-2004.
XX PF 21-JAN-2004; 2004WO-US001453.
XX PR 22-JAN-2003; 2003US-0441789P.
XX PA (UYDU-) UNIV DUKE.
XX PI Koeberl DD, Sun B;
XX WPI; 2004-571599/55.
DR P-PSDB; AD091896.
DR GENBANK; NM_000152.
XX New nucleic acid expressing lysosomal acid alpha-glucosidase (GAA)
PT polypeptide, useful for preparing a composition for treating e.g.
PT glycogen storage disease type II.
XX Claim 44; SEQ ID NO 1; 128pp; English.
XX The present invention describes an isolated nucleic acid (I) encoding a
CC chimeric polypeptide comprising a secretory signal sequence operably

CC linked to a lysosomal polypeptide. Also described: (1) a vector
CC comprising the isolated nucleic acid; (2) a pharmaceutical formulation
CC comprising the isolated nucleic acid in a carrier; (3) a cell comprising
CC the isolated nucleic acid; (4) a chimeric polypeptide comprising a
CC secretory signal sequence operably linked to a lysosomal polypeptide; (5)
CC delivering a nucleic acid encoding a lysosomal polypeptide (preferably a
CC lysosomal acid alpha-glucosidase (GAA)) to a cell; (6) producing a GAA
CC polypeptide in a cultured cell; (7) treating a deficiency of a lysosomal
CC polypeptide or lysosomal acid alpha-glucosidase (GAA) in a subject; and
CC (8) an isolated nucleic acid (II) encoding a GAA polypeptide comprising:
CC (a) a coding region encoding a GAA polypeptide; and (b) a 3' untranslated
CC region where: (i) the 3' untranslated region (UTR) comprises a GAA 3' UTR
CC comprising a deletion of at least 25 consecutive nucleotides, so that
CC upon introduction into a cell, GAA polypeptide is produced at a higher
CC level from the nucleic acid as compared with its production from a
CC nucleic acid comprising a full-length GAA 3' UTR; or (ii) the 3' UTR is
CC less than 200 nucleotides in length and comprises a segment that is
CC heterologous to the GAA coding region, so that GAA polypeptide is
CC overexpressed on introduction of the nucleic acid into a cell. (I) has
CC hepatotropic activity, and can be used in gene therapy. The nucleic acid
CC (I) encoding a chimeric polypeptide comprising a secretory signal
CC sequence operably linked to a lysosomal acid alpha-glucosidase (GAA)
CC polypeptide is useful for preparing a composition for treating lysosomal
CC acid alpha-glucosidase (GAA) deficiency e.g., glycogen storage disease
CC type II (GSD II). The present sequence encodes the human GAA amino acid
CC sequence, which is used in the exemplification of the present invention.
CC The human GAA gene is located on chromosome 17, more specifically to
CC 17q25.2-q25.3.
XX
SQ Sequence 3846 BP; 658 A; 1312 C; 1193 G; 683 T; 0 U; 0 Other;

Query Match 71.5%; Score 2647.8; DB 13; Length 3846;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2652; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Qy 1036 GAAACGGTGTGCACACCCCGCGCTCCAGAGCAGTGCACACAGTGCAGCTCCCCCCC 1095
Db 640 GATGTCACAGGCACACCCCGCGCTCCAGAGCAGTGCACACAGTGCAGCTCCCCCCC 699
Qy 1096 AACAGCGCTTCGATTGGCGCCCTGACAGAGGCATCACCAGCAAGTGCAGAGCGCGC 1155
Db 700 AACAGCGCTTCGATTGGCGCCCTGACAGAGGCATCACCAGCAAGTGCAGAGCGCGC 759
Qy 1156 GGCTGTCTACATCCCTGCAAGCAGGGGCTGCAGGGAGCCAGATGGGGCAGCCCTGG 1215
Db 760 GGCTGTCTACATCCCTGCAAGCAGGGGCTGCAGGGAGCCAGATGGGGCAGCCCTGG 819
Qy 1216 TGCTTCTTCCACCCAGCTACCCAGCTACAGCTGAGAGTGGAGAACCTGAGCTCTGAAATG 1275
Db 820 TGCTTCTTCCACCCAGCTACCCAGCTACAGCTGAGAGAACCTGAGCTCTGAAATG 879
Qy 1276 GGCTACAGGGCCACCTGACCGGTACACCCCGCTTCTCCCAAGGACATCCTGACC 1335
Db 880 GGCTACAGGGCCACCTGACCGGTACACCCCGCTTCTCCCAAGGACATCCTGACC 939
Qy 1336 CTGCGGCTGGACGCTGATGATGAGAGTGCAGAACCGCTCCACTTCACGATCAAGATCCA 1395
Db 940 CTGCGGCTGGACGCTGATGATGAGAGTGCAGAACCGCTCCACTTCACGATCAAGATCCA 999
Qy 1396 GCTAACAGCGCTACAGGTGCGCTTGGAGAGCCCGCTGTGTCAAGCGGGCAGCGTCC 1455
Db 1000 GCTAACAGCGCTACAGGTGCGCTTGGAGAGCCCGCTGTGTCAAGCGGGCAGCGTCC 1059
Qy 1456 CCACCTCTACAGGTGAGTTCCTCCAGGAGCCCTTCGGGGTGTGTCACCGGACAGTGG 1515
Db 1060 CCACCTCTACAGGTGAGTTCCTCCAGGAGCCCTTCGGGGTGTGTCACCGGACAGTGG 1119
Qy 1516 GACGGCGCGTCTGCTGAACACAGCAGCGTGGCGCCCTCTTCTTTGCGGACAGTTCCTT 1575
Db 1120 GACGGCGCGTCTGCTGAACACAGCAGCGTGGCGCCCTCTTCTTTGCGGACAGTTCCTT 1179
Qy 1576 CAGCTGTCCACCTCGCTGCTGAGTATATCAAGGCTCGCGGACACCTCAGTCC 1635

Db 1180 CAGCTGTCCACCTCGCTGCTCCTCGCAGTATATACAGAGCCTCGCCGAGCACCTCAGTCCC 1239
Qy 1636 CTGATGCTCAGACACAGCTGGACCAAGATCACCTGTGTGAAACCGGACCTTCGGCCACG 1695
Db 1240 CTGATGCTCAGACACAGCTGGACCAAGATCACCTGTGTGAAACCGGACCTTCGGCCACG 1299
Qy 1696 CCGGCTGCGAAACCTCTACCGGCTCTCACCCCTTTCTACCTGGCGCTGGAGACCGCGGCTG 1755
Db 1300 CCGGCTGCGAAACCTCTACCGGCTCTCACCCCTTTCTACCTGGCGCTGGAGACCGCGGCTG 1359
Qy 1756 GCACAGGGGTGTCTCTGCTTAACAGCAATGCCATGATGTGGTCTCGAGCGGAGCCCT 1815
Db 1360 GCACAGGGGTGTCTCTGCTTAACAGCAATGCCATGATGTGGTCTCGAGCGGAGCCCT 1419
Qy 1816 GCCCTTAGCTGAGGTGCAAGGTGGATCTCTGGATGTCTACATCTTCCTGGGCCACAG 1875
Db 1420 GCCCTTAGCTGAGGTGCAAGGTGGATCTCTGGATGTCTACATCTTCCTGGGCCACAG 1479
Qy 1876 CCACAGAGCTGGTGACAGTACCTGGACGTGTGGATACCCGTTTCATGCCGCATAC 1935
Db 1480 CCACAGAGCTGGTGACAGTACCTGGACGTGTGGATACCCGTTTCATGCCGCATAC 1539
Qy 1936 TGGGGCTGGGCTTCACCTGTGCGCTGGGCTACTCTCCACCGCTATCACCGCCAG 1995
Db 1540 TGGGGCTGGGCTTCACCTGTGCGCTGGGCTACTCTCCACCGCTATCACCGCCAG 1599
Qy 1996 GTGGTGGAAACATGACAGGGGCCCACTTCCCGCTGGACGTCCAAATGGAAACGACCTGGAC 2055
Db 1600 GTGGTGGAAACATGACAGGGGCCCACTTCCCGCTGGACGTCCAAATGGAAACGACCTGGAC 1659
Qy 2056 TACATGGACTCCGGAGGACCTTACGTTCAACAGGATGGCTTCGGGACCTTCGGGCC 2115
Db 1660 TACATGGACTCCGGAGGACCTTACGTTCAACAGGATGGCTTCGGGACCTTCGGGCC 1719
Qy 2116 ATGGTCCAGAGCTGACACAGGGCGCGCGCTATCATGATCGTGATCTTGCATC 2175
Db 1720 ATGGTCCAGAGCTGACACAGGGCGCGCGCTATCATGATCGTGATCTTGCATC 1779
Qy 2176 AGCAGCTCGGGCCCTCGCGGAGCTACAGGCGCTACAGAGGCTCTCGGAGGGGGTT 2235
Db 1780 AGCAGCTCGGGCCCTCGCGGAGCTACAGGCGCTACAGAGGCTCTCGGAGGGGGTT 1839
Qy 2236 TTCATCAACAAAGAGACCGGCGACGCTGATTTGGAAAGTATGGCCCGGGTCCACTGCC 2295
Db 1840 TTCATCAACAAAGAGACCGGCGACGCTGATTTGGAAAGTATGGCCCGGGTCCACTGCC 1899
Qy 2296 TTCCCGACTTCACCAACCCACAGCCCTGGCTGTGGAGGACATGGTGTGCTGATTC 2355
Db 1900 TTCCCGACTTCACCAACCCACAGCCCTGGCTGTGGAGGACATGGTGTGCTGATTC 1959
Qy 2356 CATGACAGGTGCTTCGACGCGCTTGTGGATTTGACATGAACAGAGCTTCCAACTTCATC 2415
Db 1960 CATGACAGGTGCTTCGACGCGCTTGTGGATTTGACATGAACAGAGCTTCCAACTTCATC 2019
Qy 2416 AGAGGCTCTGAGAGCGGCTGCCCAACATAGCTGGAGAAACCAACCTTACGTGCTGG 2475
Db 2020 AGAGGCTCTGAGAGCGGCTGCCCAACATAGCTGGAGAAACCAACCTTACGTGCTGG 2079
Qy 2476 GTGGTTGGGGGACCTTCAGGCGGCCACCATCTGTGCTCCAGCCACAGTTTCTCTCC 2535
Db 2080 GTGGTTGGGGGACCTTCAGGCGGCCACCATCTGTGCTCCAGCCACAGTTTCTCTCC 2139
Qy 2536 ACACACTAACCTGACACACTCTACGGCTGACCGAAGCCATCGCCTCCACAGGGCG 2595
Db 2140 ACACACTAACCTGACACACTCTACGGCTGACCGAAGCCATCGCCTCCACAGGGCG 2199
Qy 2596 CTGGTGAAGGCTCGGGGACACGCCCATTTGTGATCTCCCGCTCGACCTTTGCTGGCCAC 2655
Db 2200 CTGGTGAAGGCTCGGGGACACGCCCATTTGTGATCTCCCGCTCGACCTTTGCTGGCCAC 2259
Qy 2656 GGCCGATACCGCGGCCACTGGAGGGGGACGTGTGGAGCTCTCTGGAGGAGCTCGCCTCC 2715
Db 2260 GGCCGATACCGCGGCCACTGGAGGGGGACGTGTGGAGCTCTCTGGAGGAGCTCGCCTCC 2319

Qy 2716 TCCGTGCCAGAAATCCTGCAGTTTAACTGTGTGGGGTGCCTCTGTGTGGGGCCGACGTC 2775
Db 2320 TCCGTGCCAGAAATCCTGCAGTTTAACTGTGTGGGGTGCCTCTGTGTGGGGCCGACGTC 2379
Qy 2776 TCGGGCTTCTCGGGCAACACCTCAGAGGAGCTGTGTGTGGCTGAGCCAGCTGGGGGCC 2835
Db 2380 TCGGGCTTCTCGGGCAACACCTCAGAGGAGCTGTGTGTGGCTGAGCCAGCTGGGGGCC 2439
Qy 2836 TTCTACCCCTTCATGCGGAACCAACAACAGCTGTCTAGTCTGCCCCAGAGCCGTACAGC 2895
Db 2440 TTCTACCCCTTCATGCGGAACCAACAACAGCTGTCTAGTCTGCCCCAGAGCCGTACAGC 2499
Qy 2896 TTACGAGCGCGGCCACAGCCATGAGGAAGGCCCTCACCTCGCTGTACGCACTCTC 2955
Db 2500 TTACGAGCGCGGCCACAGCCATGAGGAAGGCCCTCACCTCGCTGTACGCACTCTC 2559
Qy 2956 CCCACCTCTACACACTGTTCACAGGCCACGTCGCGGGGAGACCGTGGCCCGGCC 3015
Db 2560 CCCACCTCTACACACTGTTCACAGGCCACGTCGCGGGGAGACCGTGGCCCGGCC 2619
Qy 3016 CTCTTCTGGAGTTCCCAAGGACTCTAGCAGCTGTGGAACCTGTGGACACAGCTCTGTGG 3075
Db 2620 CTCTTCTGGAGTTCCCAAGGACTCTAGCAGCTGTGGAACCTGTGGACACAGCTCTGTGG 2679
Qy 3076 GGGGAGGCCCTGCTCATCACCCAGTGTCTCAGGCCCGGGAAGGCCAAGTGACTGGCTAC 3135
Db 2680 GGGGAGGCCCTGCTCATCACCCAGTGTCTCAGGCCCGGGAAGGCCAAGTGACTGGCTAC 2739
Qy 3136 TTCCCTTGGGACATGTGTGACCTGTGAGAGCTGTGGAACCTGTGGACACAGCTCTGTGG 3195
Db 2740 TTCCCTTGGGACATGTGTGACCTGTGAGAGCTGTGGAACCTGTGGACACAGCTCTGTGG 2799
Qy 3196 CCACCCCACTGTGAGCTCCCGTGTGAGCAGCCATCCACAGCGGGGAGTGGGTGAGC 3255
Db 2800 CCACCCCACTGTGAGCTCCCGTGTGAGCAGCCATCCACAGCGGGGAGTGGGTGAGC 2859
Qy 3256 TTGGCGGCCCTCTGGAACCATCAACCTGTCCCGGCTGGGTATCATCATCCCTCTG 3315
Db 2860 TTGGCGGCCCTCTGGAACCATCAACCTGTCCCGGCTGGGTATCATCATCCCTCTG 2919
Qy 3316 CAGGCGCTGCGCTCACACACAGAGTCCCGCAGCAGCCCATGCGCCCTGCTGTGGCC 3375
Db 2920 CAGGCGCTGCGCTCACACACAGAGTCCCGCAGCAGCCCATGCGCCCTGCTGTGGCC 2979
Qy 3376 CTAAACAAAGGTTGAGAGCGCCGAGGGAGCTGTCTTGGGACGATGGAGAGCTGGAA 3435
Db 2980 CTGACCAAGGTTGAGAGCGCCGAGGGAGCTGTCTTGGGACGATGGAGAGCTGGAA 3039
Qy 3436 GTGCTGGAGCGAGGGGCCCTACACAGGTCTATCTTCTGGCCAGGAATAACAGATCGTG 3495
Db 3040 GTGCTGGAGCGAGGGGCCCTACACAGGTCTATCTTCTGGCCAGGAATAACAGATCGTG 3099
Qy 3496 AATGAGCTGTGTGCTGACAGTGTGAGAGCTGTGCTGAGAGCTGTGCTGAGAGCTGTGCT 3555
Db 3100 AATGAGCTGTGTGCTGACAGTGTGAGAGCTGTGCTGAGAGCTGTGCTGAGAGCTGTGCT 3159
Qy 3556 CTGGCGTGGCCACAGCGGCCCGCAGAGGTCTCTCTCAACGGTGTCTCTTCTTCCAACTTC 3615
Db 3160 CTGGCGTGGCCACAGCGGCCCGCAGAGGTCTCTCTCAACGGTGTCTCTTCTTCCAACTTC 3219
Qy 3616 ACCTACAGCCCGACACCAAGGTCTGTGACATCTGTGTCTGTGTGTGTGTGTGTGTGTGTGT 3675
Db 3220 ACCTACAGCCCGACACCAAGGTCTGTGACATCTGTGTCTGTGTGTGTGTGTGTGTGTGTGT 3279
Qy 3676 TTTCTCTGCTCAGCTGGTGT 3694
Db 3280 TTTCTCTGCTCAGCTGGTGT 3298

RESULT 4

ACN40294

ID ACN40294 standard; cdna; 3847 BP.

1839 TTCCATCACCAACGAGACCGCCAGCCGCTGATTGGGAAGTATGGCCCGGCTCCACTGCC 1898
2296 TTCCCGGACTTACCAACCCAGCCAGCCCTGGCTGTGGAGACATGTTGGCTGAGTTC 2355
1899 TTCCCGGACTTACCAACCCAGCCAGCCCTGGCTGTGGAGACATGTTGGCTGAGTTC 1958
2356 CATGACAGGTGCTTCGACGGCTTGTGATGACATGAACGAGGCTTCCAACTTCATC 2415
1959 CATGACAGGTGCTTCGACGGCTTGTGATGACATGAACGAGGCTTCCAACTTCATC 2018
2416 AGAGGCTCTGAGAACGGCTGCCCCAACAATGAGCTGGAGAACCCACCTTACGTCCTGG 2475
2019 AGGGGCTCTGAGAACGGCTGCCCCAACAATGAGCTGGAGAACCCACCTTACGTCCTGG 2078
2476 GTGGTTGGGGGGACCTCCAGGGGGCCACATCTGTGCTCCAGCCACAGTTTCTCTCC 2535
2079 GTGGTTGGGGGGACCTCCAGGGGGCCACATCTGTGCTCCAGCCACAGTTTCTCTCC 2138
2536 ACACACTACAACCTGACCAACCTTACGGCTGACCGAAGCCATCGCTCCACAGGGCG 2595
2139 ACACACTACAACCTGACCAACCTTACGGCTGACCGAAGCCATCGCTCCACAGGGCG 2198
2596 CTGGTGAAGGCTCGGGGACACGGCCATTTGTGATCTCCGGTTCGACCTTTGTGGCCAC 2655
2199 CTGGTGAAGGCTCGGGGACACGGCCATTTGTGATCTCCGGTTCGACCTTTGTGGCCAC 2258
2656 GGCAGATACGGCGCCACTGAGGGGGACGTGTGGAGCTCTGTGGAGCAGCTCGCTCC 2715
2259 GGCAGATACGGCGCCACTGAGGGGGACGTGTGGAGCTCTGTGGAGCAGCTCGCTCC 2318
2716 TCCGTGCCAGAACTCTGAGTTTAACTGCTGGGGGTGCTGTGGTGGGGCCAGCTC 2775
2319 TCCGTGCCAGAACTCTGAGTTTAACTGCTGGGGGTGCTGTGGTGGGGCCAGCTC 2378
2776 TGGGCTTCTGGGCAACCTCAGAGGAGCTGTGTGGTGGGACCCAGCTGGGGGCC 2835
2379 TGGGCTTCTGGGCAACCTCAGAGGAGCTGTGTGGTGGGACCCAGCTGGGGGCC 2438
2836 TTCTACCCCTTCAATGCGGACCAACACAGCTGCTCAGTCCGCCAGAGCGGTACAGC 2895
2439 TTCTACCCCTTCAATGCGGACCAACACAGCTGCTCAGTCCGCCAGAGCGGTACAGC 2498
2896 TTGAGGAGCGGGCCAGCAGGCGCATGAGAGGGCCCTCACCCTGGCTACGACTCTCTC 2955
2499 TTGAGGAGCGGGCCAGCAGGCGCATGAGAGGGCCCTCACCCTGGCTACGACTCTCTC 2558
2956 CCCACCTCTACACACTGTTTCCACAGGCGCCAGCTCGGGGGAGACCGTGGCCCGGCC 3015
2559 CCCACCTCTACACACTGTTTCCACAGGCGCCAGCTCGGGGGAGACCGTGGCCCGGCC 2618
3016 CTCTTCTGGAGTTCCCAAGGACTCTAGCACCTGAGCTGTGGACCCAGCTCTGTGG 3075
2619 CTCTTCTGGAGTTCCCAAGGACTCTAGCACCTGAGCTGTGGACCCAGCTCTGTGG 2678
3076 GGGGAGCGCCCTGTCTATCACTCCAGCTGCTCCAGGCGGGAAGCCGAACTGCTGCTAC 3135
2679 GGGGAGCGCCCTGTCTATCACTCCAGCTGCTCCAGGCGGGAAGCCGAACTGCTGCTAC 2738
3136 TTCCCTTGGGCAATGTTGACCTGACAGCGGTGCCAATAGAGCCCTTGGCAGCCTC 3195
2739 TTCCCTTGGGCAATGTTGACCTGACAGCGGTGCCAATAGAGCCCTTGGCAGCCTC 2798
3196 CCACCCCACTCTGACCTCCGTTGAGCCAGCCATCCACAGGAGGGGAGTGGGTGACG 3255
2799 CCACCCCACTCTGACCTCCGTTGAGCCAGCCATCCACAGGAGGGGAGTGGGTGACG 2858
3256 CTGCGGGCCCTCTGGACACCAATCAAGCTCCAGCTCCGGGCTGGGTACATCATCCCCCTG 3315
2859 CTGCGGGCCCTCTGGACACCAATCAAGCTCCAGCTCCGGGCTGGGTACATCATCCCCCTG 2918
3316 CAGGGCCCTGGCTCTCAACACACAGAGTCCCGCCAGCAGCCCATGCGCTGGTGGCC 3375

2919 CAGGGCCCTGGCTCACAACCAACAGAGTCCCGCCAGCAGCCCATGGCCCTGTGGCC 2978
3376 CTAAACAAAGGGTGGAGAGGCCCGAGGGAGCTGTTCTGGAGCGATGGAGAGCCTGGAA 3435
2979 CTGACCAAGGGTGGAGAGGCCCGAGGGAGCTGTTCTGGAGCGATGGAGAGCCTGGAA 3038
3436 GTGCTGGAGCGAGGGCCCTACACAGGTCATCTTCTTGGCCAGGAATAACACGATCGTG 3495
3039 GTGCTGGAGCGAGGGCCCTACACAGGTCATCTTCTTGGCCAGGAATAACACGATCGTG 3098
3496 AATGAGCTGTACGTGTGACCAAGTGAAGGAGCTGGCTGCAGCTGCAGAGGTGACTGTC 3555
3099 AATGAGCTGTACGTGTGACCAAGTGAAGGAGCTGGCTGCAGCTGCAGAGGTGACTGTC 3158
3556 CTGGGCGTGGCCACCGCGGCCCCAGCAGGTCCTCTCCAAAGGTGCTCTGTCTCCAACTTC 3615
3159 CTGGGCGTGGCCACCGCGGCCCCAGCAGGTCCTCTCCAAAGGTGCTCTGTCTCCAACTTC 3218
3616 ACCTACAGCCCGACACCAAGGTCTGTGACATCTGTGCTCTGTGATGGAGAGCAG 3675
3219 ACCTACAGCCCGACACCAAGGTCTGTGACATCTGTGCTCTGTGATGGAGAGCAG 3278
3676 TTTCTCGTCAGCTGTGTT 3694
3279 TTTCTCGTCAGCTGTGTT 3297

RESULT 5
ADP47520
ID ADF47520 standard; cDNA; 2851 BP.
XX
AC ADF47520;
XX
DT 26-FEB-2004 (first entry)
XX
DE Human GAA encoding cDNA SEQ ID NO:23.
XX
KW underglycosylated targeted therapeutic; human; lysosome;
KW lysosomal targeting domain;
KW cation-independent mannose-6-phosphate receptor;
KW underglycosylated therapeutic fusion protein; nephrotropic;
KW enzyme replacement therapy; cell therapy; gene therapy;
KW lysosomal storage disease; metabolic disease; enzyme defect;
KW Pompe disease; Tay-Sachs disease; Sandhoff disease; Fabry disease;
KW Gaucher disease; Krabbe disease; Wolman disease; Hunter disease;
KW Hunter syndrome; Sly syndrome; Schindler disease;
KW infantile sialic acid storage disease; Batten disease;
KW infantile neuronal ceroid lipofuscinosis; Ehlers-Danlos syndrome type VI;
KW congenital glycosylation disorder; gene; ss.
OS Homo sapiens.
XX
XX WO2003102583-A1.
XX
XX 11-DEC-2003.
PD
XX 29-MAY-2003; 2003WO-US017211.
XX
XX 29-MAY-2003; 2002US-0384452P.
PR 05-JUN-2002; 2002US-0386019P.
PR 06-SEP-2002; 2002US-0408816P.
PR 16-OCT-2002; 2002US-00272531.
PR 06-FEB-2003; 2003US-0445734P.
XX
XX (SYMB-) SYMBIONTICS INC.
PA
XX
PI Lebowitz JH, Beverley SM, Sly WS;
XX
XX WPI; 2004-035473/03.
DR P-PSDB; ADF47521.
XX
XX Novel underglycosylated targeted therapeutic comprising therapeutic agent
PT active in human lysosome, lysosomal targeting domain binds to human

PT	cation-independent mannose-6-phosphate receptor, for treating Pompe disease.	Db	619	CCACTCTACAGCGTGGAGTTCTCTGAGGAGCCCTTCGGGGTGATCGTGACACCGGAGCTG	678
PT		Qy	1516	GACGGCCGGTGTCTGTGAACACGACGGTGGCGCCCTGTCTTTTCGGACAGTTCTCTT	1575
XX		Db	679	GACGGCCGGTGTCTGTGAACACGACGGTGGCGCCCTGTCTTTTCGGACAGTTCTCTT	738
XX		Qy	1576	CAGCTGTCACTCTGCTGCTCGCAGTATATACAGGGCTCGCGAGACCTCTCAGTCCC	1635
CC	The present invention describes an underglycosylated targeted therapeutic (GT) which comprises: (a) a therapeutic agent that is therapeutically active in human lysosome; and (b) a lysosomal targeting domain that binds an extracellular domain of human cation-independent mannose-6-phosphate receptor (CM) and (i) does not bind a mutein in which amino acid 1572 of CM is changed from isoleucine to threonine; and (ii) binds the mutein with dissociation constant at least ten times the dissociation constant for binding CM. Also described: (i) an underglycosylated therapeutic fusion protein (FP) comprising a therapeutic domain and a subcellular targeting domain that binds to an extracellular domain of a receptor on an exterior surface of a cell, and upon internalisation of the receptor, permits localisation of the therapeutic domain to a subcellular compartment where the therapeutic domain is therapeutically active; and (2) production of GT. GT has nephrotropic activity, and can be used in enzyme replacement therapy, cell therapy and gene therapy. The FP can be used for treating a lysosomal storage disease patient by administering FP to the patient. GT can be used for treating a patient by identifying a targeting moiety that binds CM in a mannose-6-phosphate independent manner, synthesising GT comprising a therapeutic agent that is therapeutically active in a mammalian lysosome and a targeting moiety that binds CM in a mannose-6-phosphate independent manner and administering GT to the patient, where the targeting moiety is identified by screening a nucleic acid or peptide library. GT is useful for treating metabolic disease, lysosomal storage diseases and associated enzyme defects such as Pompe disease, Tay-Sachs disease, Sandhoff disease, Fabry disease, Gaucher disease, Krabbe disease, Wolman disease, Hurler syndrome, Hunter syndrome, Sly syndrome, Schindler disease, infantile sialic acid storage disease, Batten disease, infantile neuronal ceroid lipofuscinosis, Ehlers-Danlos syndrome type VI and congenital disorders of glycosylation. The present sequence is used in the exemplification of the present invention.	Db	619	CCACTCTACAGCGTGGAGTTCTCTGAGGAGCCCTTCGGGGTGATCGTGACACCGGAGCTG	678
CC		Qy	1516	GACGGCCGGTGTCTGTGAACACGACGGTGGCGCCCTGTCTTTTCGGACAGTTCTCTT	1575
CC		Db	679	GACGGCCGGTGTCTGTGAACACGACGGTGGCGCCCTGTCTTTTCGGACAGTTCTCTT	738
CC		Qy	1576	CAGCTGTCACTCTGCTGCTCGCAGTATATACAGGGCTCGCGAGACCTCTCAGTCCC	1635
CC		Db	739	CAGCTGTCCACCTCGCTGCCCTCGCAGTATATACAGGGCTCGCGAGACCTCTCAGTCCC	798
CC		Qy	1636	CTGATGCTCAGCACACAGCTGGACACAGGATCACCTGTGGAAACGGGACCTTGGCCACG	1695
CC		Db	799	CTGATGCTCAGCACACAGCTGGACACAGGATCACCTGTGGAAACGGGACCTTGGCCACG	858
CC		Qy	1696	CCCGGTGGGAACCTCTACGGGTCTCACCTTTTCTACCTGGCGCTGGAGAACGGCGGTG	1755
CC		Db	859	CCCGGTGGGAACCTCTACGGGTCTCACCTTTTCTACCTGGCGCTGGAGAACGGCGGTG	918
CC		Qy	1756	GCACAGGGGTGTTCTCTGCTAAACAGCAATGCCATGGATGTGGTCTCTGACGGAGCCCT	1815
CC		Db	919	GCACAGGGGTGTTCTCTGCTAAACAGCAATGCCATGGATGTGGTCTCTGACGGAGCCCT	978
CC		Qy	1816	GCCCTTAGCTGAGGTGCACAGGTGGGATCTCTGGATGTCTACATCTTCTGGGCCAGAG	1875
CC		Db	979	GCCCTTAGCTGAGGTGCACAGGTGGGATCTCTGGATGTCTACATCTTCTGGGCCAGAG	1038
CC		Qy	1876	CCCAAGAGCGTGGTGCAGCAGTACCTGGACGTTGTGGATACCCGTTTCATGCGGCCATAC	1935
CC		Db	1039	CCCAAGAGCGTGGTGCAGCAGTACCTGGACGTTGTGGATACCCGTTTCATGCGGCCATAC	1098
CC		Qy	1936	TGGGCGCTGGGTTCCACTGTGCGCTGGGCTACTCTTCCACCGCTATACCGGCCAG	1995
CC		Db	1099	TGGGCGCTGGGTTCCACTGTGCGCTGGGCTACTCTTCCACCGCTATACCGGCCAG	1158
CC		Qy	1996	GTGGTGGAGACATCACACAGGGCCCACTTCCCTCTGGAGTCCAAATGGAAACGACCTGGAC	2055
CC		Db	1159	GTGGTGGAGACATCACACAGGGCCCACTTCCCTCTGGAGTCCAAATGGAAACGACCTGGAC	1218
CC		Qy	2056	TACATGGAGTCCCGGAGGGACTTTCAGTTTCAACAAAGGATGGCTTCGGGACTTCGCGGC	2115
CC		Db	1219	TACATGGAGTCCCGGAGGGACTTTCAGTTTCAACAAAGGATGGCTTCGCGGACTTCGCGGC	1278
CC		Qy	2116	ATGGTGCAGGAGTGCACACAGGGCGCGCGCTACTATGATGATGATGATGATGATGATG	2175
CC		Db	1279	ATGGTGCAGGAGTGCACACAGGGCGCGCGCTACTATGATGATGATGATGATGATGATG	1338
CC		Qy	2176	AGCAGCTCGGGCCCTCGCGGAGCTACAGGCCCTACAGAGGCTTCGCGAGGGGTTCGCGAGGGGGTT	2235
CC		Db	1339	AGCAGCTCGGGCCCTCGCGGAGCTACAGGCCCTTCAGAGGCTTCGCGAGGGGTTCGCGAGGGGGTT	1398
CC		Qy	2236	TTCATCACCAACAGACCGCGCCAGCGCTGATTGGGAAAGTATGGCCCGGGTCCACTGCC	2295
CC		Db	1399	TTCATCACCAACAGACCGCGCCAGCGCTGATTGGGAAAGTATGGCCCGGGTCCACTGCC	1458
CC		Qy	2296	TTCGCCGACTTCCAAACCCCAACAGCCCTCGCTGGCTGGTGGAGGACATGTTGGTGGATTC	2355
CC		Db	1459	TTCGCCGACTTCCAAACCCCAACAGCCCTCGCTGGCTGGTGGAGGACATGTTGGTGGATTC	1518
CC		Qy	2356	CATGACAGGTCCTTTCGACGGCTTGTGGATGTGATGATGATGATGATGATGATGATGATG	2415
CC		Db	1519	CATGACAGGTCCTTTCGACGGCTTGTGGATGTGATGATGATGATGATGATGATGATGATG	1578
CC		Qy	2416	AGAGGCTCTGAGGAGCGGTGCGCCCAACCAATAGCTGGAGAACCCACCTTACCTGCTGGG	2475
CC		Db	1579	AGGGGCTCTGAGGAGCGGTGCGCCCAACCAATAGCTGGAGAACCCACCTTACCTGCTGGG	1638
CC		Qy	2476	GTGGTGGGGGACCTTCCAGCGGCCCACTTCTGTGCTCCAGCCACAGTTTCTCTCC	2535
CC		Db	1639	GTGGTGGGGGACCTTCCAGCGGCCCACTTCTGTGCTCCAGCCACAGTTTCTCTCC	1698
CC		Qy	2536	ACACATCAACCTGCACAACTCTACGGCTTCAGCGGATCGCTCCCAAGGGCG	2595
CC		Db	1699	ACACATCAACCTGCACAACTCTACGGCTTCAGCGGATCGCTCCCAAGGGCG	1758

```
QY 2596 CTGGTGAAGGCTCGGGGACACGCCCATTTGTGATCTCCCGCTCGACCTTTGCTGGCCAC 2655
Db 1759 CTGGTGAAGGCTCGGGGACACGCCCATTTGTGATCTCCCGCTCGACCTTTGCTGGCCAC 1818
QY 2656 GGCAGATACGGCGGCACCTGAGACGGGGAGTGTGTGGAGTCTCTGGAGCAGCTCGCTCC 2715
Db 1819 GGCAGATACGGCGGCACCTGAGACGGGGAGTGTGTGGAGTCTCTGGAGCAGCTCGCTCC 1878
QY 2716 TCGTGCCAGAAATCTCGAGTTTAACTGCTGGGGGTCCCTCTGCTGGGGCCGACGTC 2775
Db 1879 TCGTGCCAGAAATCTCGAGTTTAACTGCTGGGGGTCCCTCTGCTGGGGCCGACGTC 1938
QY 2776 TGGCGCTTCTTGGGCAACCTCAGAGGAGCTGTGTGTCGCTGGACCCAGCTGGGGCC 2835
Db 1939 TGGCGCTTCTTGGGCAACCTCAGAGGAGCTGTGTGTCGCTGGACCCAGCTGGGGCC 1998
QY 2836 TTCTAGCCCTTATGCGGAAACCAACAGCTGTCTCAGTCTGCTCCAGAGCGGTACAGC 2895
Db 1999 TTCTAGCCCTTATGCGGAAACCAACAGCTGTCTCAGTCTGCTCCAGAGCGGTACAGC 2058
QY 2896 TTCAGGAGCGCGCCAGCAGGCGCATGAGAGGGCCCTCACCTGCTGCTACGACTCTCTC 2955
Db 2059 TTCAGGAGCGCGCCAGCAGGCGCATGAGAGGGCCCTCACCTGCTGCTACGACTCTCTC 2118
QY 2956 CCCCACCTCTACACACTGTTCCACACAGGCCCCAGCTCGCGGGGAGACCGTGGCCCGGCC 3015
Db 2119 CCCCACCTCTACACACTGTTCCACACAGGCCCCAGCTCGCGGGGAGACCGTGGCCCGGCC 2178
QY 3016 CTCCTCTGGAGTTCCCAAGACTCTAGACCTGTGACCTGTGACACACAGCTCTGTGG 3075
Db 2179 CTCCTCTGGAGTTCCCAAGACTCTAGACCTGTGACCTGTGACACACAGCTCTGTGG 2238
QY 3076 GGGGAGCCCTGCTCATATCCCGAGTCTCCAGGCGGGAAGCCGAGTGAAGTCTGCTAC 3135
Db 2239 GGGGAGCCCTGCTCATATCCCGAGTCTCCAGGCGGGAAGCCGAGTGAAGTCTGCTAC 2298
QY 3136 TTCCCTCTGGGACATGTGACACCTGACAGCGGTGCCAATAGAGCCCTTGGCAGCCTC 3195
Db 2299 TTCCCTCTGGGACATGTGACACCTGACAGCGGTGCCAATAGAGCCCTTGGCAGCCTC 2358
QY 3196 CCACCCCACTGTGAGCTCCCGTGTAGCCAGCCATCCACAGCGGGGAGTGGGTGAGC 3255
Db 2359 CCACCCCACTGTGAGCTCCCGTGTAGCCAGCCATCCACAGCGGGGAGTGGGTGAGC 2418
QY 3256 CTGCGGCGCCCTGTGACACCATCAACGTCCACCTCCGGGCTGGGTACATCATCCCCCTG 3315
Db 2419 CTGCGGCGCCCTGTGACACCATCAACGTCCACCTCCGGGCTGGGTACATCATCCCCCTG 2478
QY 3316 CAGGGCCCTGGCCCTCAACACACAGAGTCCCGCCAGCCCATGGCCCTGGCTGTGGCC 3375
Db 2479 CAGGGCCCTGGCCCTCAACACACAGAGTCCCGCCAGCCCATGGCCCTGGCTGTGGCC 2538
QY 3376 CTAAACAAAGGTGAGAGGCGCCAGGGGAGCTGTCTGGGACGATGGAGAGAGCTGGAA 3435
Db 2539 CTAAACAAAGGTGAGAGGCGCCAGGGGAGCTGTCTGGGACGATGGAGAGAGCTGGAA 2598
QY 3436 GTGCTGAGCGAGGGGCTTACACAGGTTCATCTCTGCGCAGGAATTAACAGATCTGTG 3495
Db 2599 GTGCTGAGCGAGGGGCTTACACAGGTTCATCTCTGCGCAGGAATTAACAGATCTGTG 2658
QY 3496 AATGAGCTGGTGTAGTGTGACCAAGTGAAGGAGCTGGCTCAGCTGCAGAAAGGTGACTGT 3555
Db 2659 AATGAGCTGGTGTAGTGTGACCAAGTGAAGGAGCTGGCTCAGCTGCAGAAAGGTGACTGT 2718
QY 3556 CTGGGCGTGGCCAGCGGCGCCAGAGAGTCTCTCCAGCGGTCTCCCTGTCTCCAACTTC 3615
Db 2719 CTGGGCGTGGCCAGCGGCGCCAGAGAGTCTCTCCAGCGGTCTCCCTGTCTCCAACTTC 2778
QY 3616 ACCTACAGCCCGACACCAAGGTCCTGGAACATCTGTGCTCGCTGTGATGGAGAGCAG 3675
Db 2779 ACCTACAGCCCGACACCAAGGTCCTGGAACATCTGTGCTCGCTGTGATGGAGAGCAG 2838
```

```
QY 3676 TTTCTCGTCAGCT 3688
Db 2839 TTTCTCGTCAGCT 2851

RESULT 6
ACN42967
ID ACN42967 standard; cDNA; 3376 BP.
XX
AC ACN42967;
DT 18-NOV-2004 (first entry)
XX
DE Human diagnostic and therapeutic polynucleotide SEQ ID NO:1842.
XX
KW ss; gene; gene therapy; human diagnostic and therapeutic polynucleotide;
dithp.
XX
OS Homo sapiens.
XX
PN WO2004023973-A2.
XX
PD 25-MAR-2004.
XX
PF 12-SEP-2003; 2003WO-US028227.
XX
PR 12-SEP-2002; 2002US-0410259P.
XX
PR 12-SEP-2002; 2002US-0410260P.
XX
PA (INCY-) INCYTE CORP.
XX
PI Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
Hartshorne TA, Suchorski MT, Altus CM, Pitts SJ, Riddy LV;
Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstein EH;
Peralta CH, Anderson SP, Rioux P, Shen EJ, Wu MC, Stuve LL;
Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton BS;
Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
Patury S, Shi X, Suarez CJ;
WPI: 2004-329368/30.
P-PSDB; ABM84315.
XX
DR New diagnostic and therapeutic polynucleotides and polypeptides, useful
in diagnosing a condition, disease or disorder associated with human
molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
in gene mapping.
XX
PS Claim 1; Page; 190pp; English.
XX
CC The invention relates to novel diagnostic and therapeutic polynucleotides
selected from one of the 2722 sequences defined in the specification. A
polynucleotide of the invention may have a use in gene therapy. The human
diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
used to diagnose a particular condition, disease or disorder associated
with human molecules, e.g. cell proliferative disorders,
autoimmune/inflammatory disorder, developmental disorder, endocrine
disorder, neurological disorders, gastrointestinal disorders, or
infections caused by virus, bacteria, fungi or parasite. The dithp
molecules may also be used in genetic mapping, in identifying individuals
from minute biological samples, in detecting single nucleotide
polymorphisms, as molecular weight markers, and for somatic or germline
gene therapy. The present sequence represents a dithp polynucleotide of
the invention. Note: The sequence data for this patent is not represented
in the printed specification, but was obtained in electronic format
directly from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX
SQ Sequence 3376 BP; 590 A; 1153 C; 1020 G; 613 T; 0 U; 0 Other;
Query Match 59.1%; Score 2186.2; DB 13; Length 3376;
Best Local Similarity 91.2%; Pred. No. 0;
Matches 2425; Conservative 0; Mismatches 13; Indels 221; Gaps 1;
```

1036 GAAACCGGTGCACACCCCGGCGTCCAGAGCAAGTGCACACAGTGCAGAGTCCCGCC 1095
1096 AACAGCGCTTCGATTGCGCCCTGACAGAGCCATCACAGGAAACAGTGCAGAGCGCGC 1155
514 AACAGCGCTTCGATTGCGCCCTGACAGAGCCATCACAGGAAACAGTGCAGAGCGCGC 573
1156 GGTGCTGTCTACATCCCTGCAAGAGCGGCTGACAGGAGCCAGATGAGGAGCCCTG 1215
574 GGTGCTGTCTACATCCCTGCAAGAGCGGCTGACAGGAGCCAGATGAGGAGCCCTG 633
1216 TGTCTTCTCCACCGAGTACCCAGCTACAGCTGAGAACTGAGCTCTCTGAATG 1275
634 TGTCTTCTCCACCGAGTACCCAGCTACAGCTGAGAACTGAGCTCTCTGAATG 693
1276 GGTACAGCGCCACCTGACCCCTGACACCCCACTCTTCTCCCAAGGACATCTGACC 1335
694 GGTACAGCGCCACCTGACCCCTGACACCCCACTCTTCTCCCAAGGACATCTGACC 753
1336 CTGCGCTGGAAGTGTGATGAGAGTGAAGACCGCTCCACTTCAAGTCAAGATCCA 1395
754 CTGCGCTGGAAGTGTGATGAGAGTGAAGACCGCTCCACTTCAAGTCAAGATCCA 813
1396 GCTAACAGCGGCTACAGAGTGCCTTGGAGACCCCGCTGTCAAGCCGAGCACCGTCC 1455
814 GCTAACAGCGGCTACAGAGTGCCTTGGAGACCCCGCTGTCAAGCCGAGCACCGTCC 873
1456 CCACTTACAGCGTGAAGTTCCTCGAGGAGCCCTCGGGGTGATCTGACACCGGAGCTG 1515
874 CCACTTACAGCGTGAAGTTCCTCGAGGAGCCCTCGGGGTGATCTGACACCGGAGCTG 933
1516 GACGGCGGGTGTGTGTAACAGAGCGTGGCGCCCTGTCTTTCGGACCAAGTTCCTT 1575
934 GACGGCGGGTGTGTGTAACAGAGCGTGGCGCCCTGTCTTTCGGACCAAGTTCCTT 993
1576 CAGCTGTCAAGTGTGCTGCTGCAAGAGTATACAGAGCTGCGGAGCACTCAGTCC 1635
994 CAGCTGTCAAGTGTGCTGCTGCAAGAGTATACAGAGCTGCGGAGCACTCAGTCC 1053
1636 CTGATGCTCAGCAACAGCTGGAACAGATCACCTGTGGAACCGGACCTTGGCGCCAG 1695
1054 CTGATGCTCAGCAACAGCTGGAACAGATCACCTGTGGAACCGGACCTTGGCGCCAG 1113
1696 CCGGTGCGAACTCTACGGGTCTACACCTTCTACCTGGCGTGAAGACCGCGGCTG 1755
1114 CCGGTGCGAACTCTACGGGTCTACACCTTCTACCTGGCGTGAAGACCGCGGCTG 1173
1756 GCACAGGGGTGTCTGTCTAAACAGCAATGCAATGCAATGCAATGCAATGCAATGCA 1815
1174 GCACAGGGGTGTCTGTCTAAACAGCAATGCAATGCAATGCAATGCAATGCAATGCA 1233
1816 GCCCTTAGCTGGAGTGCAGGTGGATGCTGATGCTACATCTTCTGGGCGCAGAG 1875
1234 GCCCTTAGCTGGAGTGCAGGTGGATGCTGATGCTACATCTTCTGGGCGCAGAG 1293
1876 CCCAAGAGCGTGTGCAGAGTACCTGAGCTTGTGGATACCCGTTCAATGCGCCATAC 1935
1294 CCCAAGAGCGTGTGCAGAGTACCTGAGCTTGTGGATACCCGTTCAATGCGCCATAC 1353
1936 TGGGGCTTGGGCTTCACTGTGCGCTGGGCTACTCTCTCAACCGCTATCAACCGCCAG 1995
1354 TGGGGCTTGGGCTTCACTGTGCGCTGGGCTACTCTCTCAACCGCTATCAACCGCCAG 1413
1996 GTGGTGGAGAACTGACAGGGGCTTCTCCCGCTGGAAGTCAATGGAACGACCTGGAC 2055
1414 GTGGTGGAGAACTGACAGGGGCTTCTCCCGCTGGAAGTCAATGGAACGACCTGGAC 1473
2056 TACATGAGCTCCCGAGGAGCTTCAAGTTCACCAAGAGTGGCTTCCGGGACTTCCCGGC 2115
1474 TACATGAGCTCCCGAGGAGCTTCAAGTTCACCAAGAGTGGCTTCCGGGACTTCCCGGC 1533
2116 ATGGTGCAGGAGCTGCAACAGGGCGCGCGCTACATGATGATGATGATGATGATGATGAT 2175

1534 ATGGTGCAGGAGCTGCAACAGGGCGCGCTACATGATGATGATGATGATGATGATGATGAT 1593
2176 AGCAGCTCGGGCCCTGCGGGAGCTACAGCCCTACGACAGGGTCTGCGGAGGGGGTT 2235
1594 AGCAGCTCGGGCCCTGCGGGAGCTACAGCCCTACGACAGGGTCTGCGGAGGGGGTT 1653
2236 TTTCATCAACAAAGAGACCGGCTGATGAGAGGATGAGAGGATGAGAGGATGAGAGGATGAG 2295
1654 TTTCATCAACAAAGAGACCGGCTGATGAGAGGATGAGAGGATGAGAGGATGAGAGGATGAG 1713
2296 TTTCATCAACAAAGAGACCGGCTGATGAGAGGATGAGAGGATGAGAGGATGAGAGGATGAG 2355
1714 TTTCATCAACAAAGAGACCGGCTGATGAGAGGATGAGAGGATGAGAGGATGAGAGGATGAG 1773
2356 CATGACAGGCTCGCCCTTTCAGCGCTTGTGGAATGATGATGATGATGATGATGATGATGATGAT 2415
1774 CATGACAGGCTCGCCCTTTCAGCGCTTGTGGAATGATGATGATGATGATGATGATGATGATGAT 1833
2416 AGAGGCTTGAGGACCGCTGCGCCCAACATGATGATGAGAGGATGAGAGGATGAGAGGATGAG 2475
1834 AGGGGCTTGAGGACCGCTGCGCCCAACATGATGATGAGAGGATGAGAGGATGAGAGGATGAG 1893
2476 GTGGTGGGGGACCGCTCCAGGGCGCCACATCTGTGCTCCAGCAGGATGATGATGATGATGAT 2535
1894 GTGGTGGGGGACCGCTCCAGGGCGCCACATCTGTGCTCCAGCAGGATGATGATGATGATGAT 1953
2536 ACACACTACAACTGCAACCTCTACCGCTGACCGAGGATGATGATGATGATGATGATGATGATGAT 2595
1954 ACACACTACAACTGCAACCTCTACCGCTGACCGAGGATGATGATGATGATGATGATGATGATGAT 2013
2596 CTGGTGAAGGCTCGGGGACCGCCATTTGTGATCTCCCGCTGAGACCTTTGTGCTGCGCAC 2655
2014 CTGGTGAAGGCTCGGGGACCGCCATTTGTGATCTCCCGCTGAGACCTTTGTGCTGCGCAC 2073
2656 GGGGATACCGCGGCACTGGAGCGGGGACGTTGGAGCTCTGGAGAGAGCTCGGCTCC 2715
2074 GGGGATACCGCGGCACTGGAGCGGGGACGTTGGAGCTCTGGAGAGAGCTCGGCTCC 2133
2716 TCGGTGCAAGAACTCTGCAAGTTTAACTGCTGGGGGCTGCTTGTGCTGGGGCGGACGCTC 2775
2134 TCGGTGCAAGAACTCTGCAAGTTTAACTGCTGGGGGCTGCTTGTGCTGGGGCGGACGCTC 2193
2776 TCGGCTTCTCGGGCAACCTCAGAGGAGCTGTGTGCTGCTGAGACCGAGCTGGGGGCTC 2835
2194 TCGGCTTCTCGGGCAACCTCAGAGGAGCTGTGTGCTGCTGAGACCGAGCTGGGGGCTC 2253
2836 TTCTACCGCTTCTGCGGAAACCAACAGCTGCTCAGTCTGCTCCCGAGAGCGGTACAGC 2895
2254 TTCTACCGCTTCTGCGGAAACCAACAGCTGCTCAGTCTGCTCCCGAGAGCGGTACAGC 2313
2896 TTCTACCGCTTCTGCGGAAACCAACAGCTGCTCAGTCTGCTCCCGAGAGCGGTACAGC 2955
2314 TTCTACCGCTTCTGCGGAAACCAACAGCTGCTCAGTCTGCTCCCGAGAGCGGTACAGC 2373
2956 CCCACCTCTACACACTGTTCACAGGCGGCTGCTGCGGGGAGAGCGGTGCGCGCGGCTC 3015
2374 CCCACCTCTACACACTGTTCACAGGCGGCTGCTGCGGGGAGAGCGGTGCGCGCGGCTC 2433
3016 CTCCTCTGGAGTTCGCCAAGGACTTAGACCTGAGCTGAGTGTGGAACCAACAGCTCTGTGG 3075
2434 CTCCTCTGGAGTTCGCCAAGGACTTAGACCTGAGCTGAGTGTGGAACCAACAGCTCTGTGG 2470
3076 GGGGAGGCGCTCTCATCACCCAGTGTCTCAGGCGGGAGAGCGGAGTGTGCTGCTAC 3135
2471 ----- 2470
3136 TTCCCTTTGGGCACTGTGTACGACCTGAGAGCGGTGCGCAATAGAGGCGCTTGGAGCGCTC 3195
2471 ----- 2470
3196 CCACCCCGACCTGACGCTCCCGTGGAGCGGAGCCATCCACAGCGAGGGGAGTGGGTGACG 3255

```
Db 2471 ----- 2470
QY 3256 CTGCCGCCCCCTGGACACCACTCAACGTCACCTCCCGGCTGGGTACATCATCCCCCTGG 3315
Db 2471 -----ACCATCAAGTCTCACTCCGGGCTGGGTACATCATCCCCCTGG 2512
QY 3316 CAGGGCCCTGGCTCAAAACCAAGAGTCCCGCCAGCAGCCCATGCTGTGGCC 3375
Db 2513 CAGGGCCCTGGCTCAAAACCAAGAGTCCCGCCAGCAGCCCATGCTGTGGCC 2572
QY 3376 CTAACCAAGGTGGAGAGCCCGAGGGAGCTGTTCTGGAGATGAGAGAGCTGGNA 3435
Db 2573 CTGACCAAGGGTGGAGAGCCCGAGGGAGCTGTTCTGGAGATGAGAGAGCTGGAA 2632
QY 3436 GTGCTGGAGAGGGGCTTACACACAGGTCTCTTCTGGCCAGGAATAACACGATCGTG 3495
Db 2633 GTGCTGGAGAGGGGCTTACACACAGGTCTCTTCTGGCCAGGAATAACACGATCGTG 2692
QY 3496 AATGAGCTGGTACGTGTGACCAAGTGAAGGAGCTGGCTTCAGCTGCAGAAAGTGTGTC 3555
Db 2693 AATGAGCTGGTACGTGTGACCAAGTGAAGGAGCTGGCTTCAGCTGCAGAAAGTGTGTC 2752
QY 3556 CTGGGCGTGGCCACGGCGCCCGACGAGTCTCTTCCAAAGGTGTCCCTGTCTCCAACTTC 3615
Db 2753 CTGGGCGTGGCCACGGCGCCCGACGAGTCTCTTCCAAAGGTGTCCCTGTCTCCAACTTC 2812
QY 3616 ACCTACAGCCCGACACCAAGGTCCTGGACATCTGTGTCGTGTGATGGAGAGCAG 3675
Db 2813 ACCTACAGCCCGACACCAAGGTCCTGGACATCTGTGTCGTGTGATGGAGAGCAG 2872
QY 3676 TTTCTCGTCAGCTGGTGT 3694
Db 2873 TTTCTCGTCAGCTGGTGT 2891

RESULT 7
ACN42966
ID ACN42966 standard; cDNA; 3455 BP.
AC ACN42966;
XX ACN42966;
DT 18-NOV-2004 (first entry)
DE Human diagnostic and therapeutic polynucleotide SEQ ID NO:1841.
XX
KW es; gene; gene therapy; human diagnostic and therapeutic polynucleotide;
KW dithp.
XX Homo sapiens.
XX WO2004023973-A2.
XX 25-MAR-2004.
XX
XX 12-SEP-2003; 2003WO-US028227.
XX
XX 12-SEP-2002; 2002US-0410259P.
XX
XX 12-SEP-2002; 2002US-0410260P.
XX
XX (INCY-) INCYTE CORP.
XX
XX Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
XX Harthshorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;
XX Mooney EM, Deleagane AM, Panesar IS, Banville SC, Reddy TP;
XX Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;
XX Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
XX Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kilton ES;
XX Xu Y, Kwong M, Pollicky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
XX Patury S, Shi X, Suarez CJ;
XX
XX WPI: 2004-329368/30.
XX P-PSDB; ABM84314.
XX
```

```
PT New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT in diagnosing a condition, disease or disorder associated with human
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT in gene mapping.
XX Claim 1; Page; 190pp; English.
XX The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human
CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorders, developmental disorder, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp polynucleotide of
CC the invention. Note: The sequence data for this patent is not represented
CC in the printed specification, but was obtained in electronic format
CC directly from WIFO at www.wipo.int/pct/en/sequences/listing.htm
XX Sequence 3455 BP; 597 A; 1189 C; 1042 G; 627 T; 0 U; 0 Other;
SQ Query Match 56.7%; Score 2097.2; DB 13; Length 3455;
Best Local Similarity 88.6%; Pred. No. 0;
Matches 2425; Conservative 0; Mismatches 13; Indels 300; Gaps 2;
QY 1036 GAAACCGGTGACACCCCGGCGTCCAGAGCAGTGGCCACACAGTGGACGTCGCCCC 1095
Db 454 GATGCCAGGACACCCCGGCGTCCAGAGCAGTGGCCACACAGTGGACGTCGCCCC 513
QY 1096 TACAGCCGCTTCGATTGGCGCCCTGCACAGGCCATCACCCAGGAACAGTGGCAGGCCCG 1155
Db 514 AACAGCCGCTTCGATTGGCGCCCTGCACAGGCCATCACCCAGGAACAGTGGCAGGCCCG 573
QY 1156 GGCTGCTGTACATCCCTGCAAGCAGGGGGTGCAGGGAGCCAGATGGGGCAGCCCTGG 1215
Db 574 GGCTGCTGTACATCCCTGCAAGCAGGGGGTGCAGGGAGCCAGATGGGGCAGCCCTGG 633
QY 1216 TGCTTTCTCCACCAGCTACCCAGCTACAGCTGAGAGAACCTGAGTCTCTGAATG 1275
Db 634 TGCTTTCTCCACCAGCTACCCAGCTACAGCTGAGAGAACCTGAGTCTCTGAATG 693
QY 1276 GGCTACAGGGCCACCTGCACCGTACCCACCCCTCTTCCCAAGACATCTCTGACC 1335
Db 694 GGCTACAGGGCCACCTGCACCGTACCCACCCCTCTTCCCAAGACATCTCTGACC 753
QY 1336 CTGCGGCTGGAGCTGATGAGAGTGGAGAACCGCCCTCCACTTCACGATCAAGATCCA 1395
Db 754 CTGCGGCTGGAGCTGATGAGAGTGGAGAACCGCCCTCCACTTCACGATCAAGATCCA 813
QY 1396 GCTAACAGCGCTACGAGGTGCTTGGAGAGCCCGCGTGTCCAGCGGGCAGCGTCC 1455
Db 814 GCTAACAGCGCTACGAGGTGCTTGGAGAGCCCGCGTGTCCAGCGGGCAGCGTCC 873
QY 1456 CCACCTACAGCGTGGAGTTCCTCGAGGAGCCCTTCGGGGTGTATCTGCGCCGCGAGCTG 1515
Db 874 CCACCTACAGCGTGGAGTTCCTCGAGGAGCCCTTCGGGGTGTATCTGCGCCGCGAGCTG 933
QY 1516 GACGGCCGCTGCTGCTGAACACAGCAGGTGGCGCCCTGTCTTTTGGAGACAGTTCCTT 1575
Db 934 GACGGCCGCTGCTGCTGAACACAGCAGGTGGCGCCCTGTCTTTTGGAGACAGTTCCTT 993
QY 1576 CAGCTGTCCACCTCGCTGCTGAGTATATACAGGGCTGCGGAGACCTCAGTCCC 1635
Db 994 CAGCTGTCCACCTCGCTGCTGAGTATATACAGGGCTGCGGAGACCTCAGTCCC 1053
QY 1636 CTGATGCTCAGCAGCTGGACAGGATCACCTCTGTGAACCGGAGCTTCGGCCACG 1695
Db 1054 CTGATGCTCAGCAGCTGGACAGGATCACCTCTGTGAACCGGAGCTTCGGCCACG 1113
```

1696 CCGGTGCGAACTCTACGGGTCTCAACCTTTTCTACTGCGCTGAGAGACGCGGGTCG 1755
1114 CCGGTGCGAACTCTACGGGTCTCAACCTTTTCTACTGCGCTGAGAGACGCGGGTCG 1173
1756 GCACGGGGTGTCTCTTAACACGAAATCCATGGATGTGCTCTGACGCCGACCT 1815
1174 GCACGGGGTGTCTCTTAACACGAAATCCATGGATGTGCTCTGACGCCGACCT 1233
1816 GCCCTTAGCTGAGGTGACAGGTGGATCCTCGATGTCTACATCTTCTGGGCCACAG 1875
1234 GCCCTTAGCTGAGGTGACAGGTGGATCCTCGATGTCTACATCTTCTGGGCCACAG 1293
1876 CCAGAGCGTGTGACGAGTACTTGGACGTTGTG----- 1911
1294 CCAGAGCGTGTGACGAGTACTTGGACGTTGTGAGGCTGCTCCCTGGCGCG 1353
1912 ----- 1916
1354 GCGCCGCCCCAAGGTCCCTCTCCCTCCCTCATGAAGTCGGCGTTGGCTGACAGATA 1413
1917 CCGGTTATGCGCCCATACTGGGGCTTGGGCTTCCACCTGTGCGCTGGGGCTACTCTC 1976
1414 CCGGTTATGCGCCCATACTGGGGCTTGGGCTTCCACCTGTGCGCTGGGGCTACTCTC 1473
1977 CACCGTATCACCGCCAGGTGTGAGAAATGACACAGGGCCACTTCCCTCGGACGT 2036
1474 CACCGTATCACCGCCAGGTGTGAGAAATGACACAGGGCCACTTCCCTCGGACGT 1533
2037 CCATGGAACGACTGACATGACATCCCGAGGGACTTACGTTTCAAGGATGG 2096
1534 CCATGGAACGACTGACATGACATCCCGAGGGACTTACGTTTCAAGGATGG 1593
2097 CTTCCGGGACTTCCCGCCATGTGAGAGCTGACACAGGGCGCGGCTACATGAT 2156
1594 CTTCCGGGACTTCCCGCCATGTGAGAGCTGACACAGGGCGCGGCTACATGAT 1653
2157 GATGTGGATCTGCCATCAGCAGCTCGGGCCCTGCGGAGCTACAGGCCCTACGACGA 2216
1654 GATGTGGATCTGCCATCAGCAGCTCGGGCCCTGCGGAGCTACAGGCCCTACGACGA 1713
2217 GGGTCTCGGAGGGGGTTCATACCAACAGAGACGGGCCAGCCCTGATGGGAGGT 2276
1714 GGGTCTCGGAGGGGGTTCATACCAACAGAGACGGGCCAGCCCTGATGGGAGGT 1773
2277 ATGGCCGGGTCCACTGCTCCCGACTTCCACAAACAGAGCTGAGCCCTGGTGGGA 2336
1774 ATGGCCGGGTCCACTGCTCCCGACTTCCACAAACAGAGCTGAGCCCTGGTGGGA 1833
2337 GGAATGCTGGCTGAGTTCCATGACAGGTGCCCTTCGACGGCTTGGATTGACATGAA 2396
1834 GGAATGCTGGCTGAGTTCCATGACAGGTGCCCTTCGACGGCTTGGATTGACATGAA 1893
2397 CGAGCTTCCAACTTCAACAGGCTCTGAGGACGGCTGCCCAACATGAGCTGAGAA 2456
1894 CGAGCTTCCAACTTCAACAGGCTCTGAGGACGGCTGCCCAACATGAGCTGAGAA 1953
2457 CCACCTTACGCTGCTGGGGTGGGGGACCTCCAGGGCGGCCACCATCTGTCCTC 2516
1954 CCACCTTACGCTGCTGGGGTGGGGGACCTCCAGGGCGGCCACCATCTGTCCTC 2013
2517 CAGCCACCAAGTTTCTCTCACACATACAACTTGCAACCTCTAGGGCTGACCGAAGC 2576
2014 CAGCCACCAAGTTTCTCTCACACATACAACTTGCAACCTCTAGGGCTGACCGAAGC 2073
2577 CATCGCTCCCAACAGGGCTGTGAGGCTCGGGGACACGCCCATTTGTGATCTCCG 2636
2074 CATCGCTCCCAACAGGGCTGTGAGGCTCGGGGACACGCCCATTTGTGATCTCCG 2133
2637 CTCGACTTGTGGCCACGGCCGATACGCCGGCCACTGACGGGGAGCTGTGGAGCTC 2696
2134 CTCGACTTGTGGCCACGGCCGATACGCCGGCCACTGACGGGGAGCTGTGGAGCTC 2193

RESULT 8
ACN42968
ID ACN42968 standard; cDNA; 3105 BP.
XX

2697 CTGGAGAGAGCTCGCTCCTCCGTGCCAGAAATCTGCAAGTTTAACTGCTGGGGGTGCC 2756
2194 CTGGAGAGAGCTCGCTCCTCCGTGCCAGAAATCTGCAAGTTTAACTGCTGGGGGTGCC 2253
2757 TCTGGTCCGGGCGGAGCTGCTGGGCTTCTGGGCAACACCTCAGAGGAGCTGTGTGTGCG 2816
2254 TCTGGTCCGGGCGGAGCTGCTGGGCTTCTGGGCAACACCTCAGAGGAGCTGTGTGTGCG 2313
2817 CTGGACCCAGCTGGGGGCTTCTTACCCCTTATGCGGAAACCAACAGAGCTGCTCAGTCT 2876
2314 CTGGACCCAGCTGGGGGCTTCTTACCCCTTATGCGGAAACCAACAGAGCTGCTCAGTCT 2373
2877 GCGCCAGGAGCGGTACAGCTTCAGGAGCGCGCCAGAGGCGCATAGAGAGCCCTCAC 2936
2374 GCGCCAGGAGCGGTACAGCTTCAGGAGCGCGCCAGAGGCGCATAGAGAGCCCTCAC 2433
2937 CTTGGCTACGCACTCTCCCGCCACCTCTACACATGTTTCCACAGGCGCCAGCTCGCGG 2996
2434 CTTGGCTACGCACTCTCCCGCCACCTCTACACATGTTTCCACAGGCGCCAGCTCGCGG 2493
2997 GGAGACCGTGGGCCCGGCTTCTTCTGGAGTTCCTCAAGGACTCTAGCACCTGGAAGTGT 3056
2494 GGAGACCGTGGGCCCGGCTTCTTCTGGAGTTCCTCAAGGACTCTAGCACCTGGA--- 2549
3057 GGACCAACAGCTCTGTGGGGGAGGCGCTGCTCATCACCCAGTCTCCAGGCGCGGAA 3116
2550 ----- 2549
3117 GCGCCAGTGAAGTGGTACTTCCCTTGGGACATGTTGACAGCTTCAGAGCGGTGCCAAT 3176
2550 ----- 2549
3177 AGAGGCGCTTGGCAGCTTCCACCCAGCTCGAGCTCCCGGTGAGCCAGCATCCACAG 3236
2550 ----- 2549
3237 CGAGGGGAGTGGGTGACGCTCGCGGCCCCCTGGACACCATCAAGTCCACCTCCGGGC 3296
2550 ----- ACCATCAAGTCCACCTCCGGGC 2572
3297 TGGGTACATCATCCCCCTGACGGGCTTGGGCTTCAACACAGAGTCCCGCCAGAGCC 3356
2573 TGGGTACATCATCCCCCTGACGGGCTTGGGCTTCAACACAGAGTCCCGCCAGAGCC 2532
3357 CATGGCCCTGGCTGGGCTTAAACCAAGGCTGGAGAGGCGCCGAGGGGAGCTGTTCTGGGA 3416
2633 CATGGCCCTGGCTGGGCTTAAACCAAGGCTGGAGAGGCGCCGAGGGGAGCTGTTCTGGGA 2592
3417 CGATGGAGAGAGCTGGAAAGTCTGGAGCGGAGGCGCTTACACAGAGTCACTTCTCTGCG 3476
2693 CGATGGAGAGAGCTGGAAAGTCTGGAGCGGAGGCGCTTACACAGAGTCACTTCTCTGCG 2752
3477 CAGGAATACACAGATCGTAATGCTGTGACCTGTGACAGTGGAGGAGCTGGCCCTGCA 3536
2753 CAGGAATACACAGATCGTAATGCTGTGACCTGTGACAGTGGAGGAGCTGGCCCTGCA 2812
3537 GCTGCAGAGAGTGAAGTCTGCTGGGGTGGCCACGGCGCCAGAGAGTCTCTCTCAACGG 3596
2813 GCTGCAGAGAGTGAAGTCTGCTGGGGTGGCCACGGCGCCAGAGAGTCTCTCTCAACGG 2872
3597 TGTCCCTGTCTCCAACTTACCTACAGCCCGGACACCAAGGCTCCTGGACATCTGTGTCTC 3656
2873 TGTCCCTGTCTCCAACTTACCTACAGCCCGGACACCAAGGCTCCTGGACATCTGTGTCTC 2932
3657 GCTGTGTGAGAGAGAGTCTGCTGCTCAGCTGGTGT 3694
2933 GCTGTGTGAGAGAGAGTCTGCTGCTCAGCTGGTGT 2970

Db 1654 TTCTATCAACAAAGAGACCGGCGCGCTGATTGGAGGATATGGCCCGGTCACTGCC 1713
Qy 2296 TTCCCGGACTTACCAACCCACAGCCCTGGCTGGTGGAGGACATGGTGGCTGATTC 2355
Db 1714 TTCCCGGACTTACCAACCCACAGCCCTGGCTGGTGGAGGACATGGTGGCTGATTC 1773
Qy 2356 CATGACAGGTGGCTTCGAGCGGTTGTGATGACATGACATGACAGCCCTTCCAACTTCATC 2415
Db 1774 CATGACAGGTGGCTTCGAGCGGTTGTGATGACATGACATGACAGCCCTTCCAACTTCATC 1833
Qy 2416 AGAGGCTCTGAGGAGCGCTGCCCAACAATGAGCTGGAGAACCCACCTACGTGCCTGGG 2475
Db 1834 AGGGGCTCTGAGGAGCGCTGCCCAACAATGAGCTGGAGAACCCACCTACGTGCCTGGG 1893
Qy 2476 GTGGTTGGGGGACCTTCCAGGGGCGCCACCATCTGTGCTTCCAGCCACCAAGTTTCTCTCC 2535
Db 1894 GTGGTTGGGGGACCTTCCAGGGGCGCCACCATCTGTGCTTCCAGCCACCAAGTTTCTCTCC 1953
Qy 2536 ACACACTACAACTGACAACTCTAGGGCTGACCGAGCCATCGCTCCACAGGGCG 2595
Db 1954 ACACACTACAACTGACAACTCTAGGGCTGACCGAGCCATCGCTCCACAGGGCG 2013
Qy 2596 CTGCTGAAGGCTCGGGGACACGCCATTGTGATCTCCCGCTCGACCTTTGCTGGCCAC 2655
Db 2014 CTGCTGAAGGCTCGGGGACACGCCATTGTGATCTCCCGCTCGACCTTTGCTGGCCAC 2073
Qy 2656 GGCCTGATACCGCGGCGCACTGGAGCGGGGACGTGTGGAGCTCTCTGGGAGCAGCTCGCTCC 2715
Db 2074 GGCCTGATACCGCGGCGCACTGGAGCGGGGACGTGTGGAGCTCTCTGGGAGCAGCTCGCTCC 2133
Qy 2716 TCCGTGCGAGAAATCTGCAAGTTAACTCTGCTGGGGTGCTCTGTGCGGGCGGACGTC 2775
Db 2134 TCCGTGCGAGAAATCTGCAAGTTAACTCTGCTGGGGTGCTCTGTGCGGGCGGACGTC 2193
Qy 2776 TCGGGCTTCTTGGGCAACCTCAGAGGAGCTGTGTGGCTGGACCCAGCTGGGGGCC 2835
Db 2194 TCGGGCTTCTTGGGCAACCTCAGAGGAGCTGTGTGGCTGGACCCAGCTGGGGGCC 2253
Qy 2836 TTCTACCCCTTTCATGGGGAAACCAACAGCTCTCTCAGTCTGCGCCAGAGCGCTACAGC 2895
Db 2254 TTCTACCCCTTTCATGGGGAAACCAACAGCTCTCTCAGTCTGCGCCAGAGCGCTACAGC 2313
Qy 2896 TTCAGCGAGCGCGCCAGCAGGCGCATGAGGAAGGCCCTACCCCTG 2940
Db 2314 TTCAGCGAGCGCGCCAGCAGGCGCATGAGGAAGGCCCTACCCCTG 2358

RESULT 9

ABI99468
ID ABI99468 standard; cDNA; 3364 BP.

AC ABI99468;

DT 07-MAR-2002 (first entry)

DE Mouse ischaemic condition related cDNA sequence SEQ ID NO:421.

XX Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;

KW vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.

XX Mus musculus.

XX WO200188188-A2.

XX 22-NOV-2001.

XX 18-MAY-2001; 2001WO-0145977.

XX 18-MAY-2000; 2000JP-00145977.

XX (UYN1-) UNIV NIHON SCHOOL JURIDICAL PERSON.

XX Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;

Qy 1648 ACCAGCTGGACCGAGGATCAACCTGTGGAAACCGGAGACCTTTGCGCCCAACGCCCGGTGCGAAC 1707

XX WPI; 2002-034733/04.
DR P-PSDB; ABB57174.

XX Examining the ischemic condition (e.g. occlusive ischemia) by measuring
PT expression levels of particular genes defined in the specification or by
PT determining the expression profile of a gene group comprising these
XX genes.

XX Claim 2; Page 1147-1153; 2690pp; English.

XX The present invention describes a method for examining ischaemic
CC conditions, comprising measuring the expression levels of particular
CC genes (1) in a test sample or determining the expression profile of a
CC gene group in the sample comprising genes selected from (1). The method
CC is useful for examining the ischaemic condition (e.g. compressive
CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring
CC expression levels of particular genes (ABI99202 to ABI99912, encoding the
CC protein sequences in ABB57020 to ABB57374) or by determining the
CC expression profile of a gene group comprising these genes. The expression
CC levels or expression profiles produced by these genes are used as an
CC indicator when screening for ischaemic condition-improving drugs or
CC therapeutics for ischaemic diseases. ABI99913 and ABI99914 represent PCR
CC primers for a mouse ischaemic condition related sequence, which are used
CC in the exemplification of the present invention

XX Sequence 3364 BP; 695 A; 1057 C; 918 G; 692 T; 0 U; 2 Other;

Qy Query Match 50.8%; Score 1881.4; DB 6; Length 3364;

Db Best Local Similarity 82.1%; Pred. No. 0;

Qy Matches 2174; Conservative 2; Mismatches 468; Indels 3; Gaps 1;

Qy 1048 CACCCCGCGCGTCCAGAGCAGTGCACACAGTGCAGCTGCCCGCCCAACAGCCGCTTC 1107

Db 323 CAGACTGAACAGCCCAAGAACCCACACAGTGTGATGTGCCCGCCAGCAGCCGCTTT 382

Qy 1108 GATTGCGCCCTTGACAAAGCCATCACCCAGGAACAGTGGAGAGCCCGCGGCTCTGTGTAC 1167

Db 383 GACTGTGCCCGCCGACAAAGGCATCTCACAGGAGCAATGGAGCCCGAGGCTGTGTAT 442

Qy 1168 ATCCCTGCAACAGCGGCTGCAGGAGCCCGAGATGGGCGCCGCTGTCTTCCCA 1227

Db 443 GTCCAGCAGCGGCGGTGTGAAGGAGCCACAGATAGGGGAGCCCTGTGTCTTCCCT 502

Qy 1228 CCAGCTACCCAGCTTACAAAGCTGGAGAACCTTGAGTCTCTCTGAAATGGCTACAGGCC 1287

Db 503 CCAGCTACCCAGCTACCGTCTAGAGAACCTTGAGTCTTACAGAGTCGGGTACACAGCC 562

Qy 1288 ACCCTGACCCGTACACCCCGACCTTTTCCCAAGGACATCTGTGACCTGCGGGTGGAC 1347

Db 563 ACCCTGACCCGTACACCGCCGACCTTCTTCCCAAGGATGTGTGACCTTACAGCTGGAG 622

Qy 1348 GTGATGATGGAGACTGAGAACCGCTTCCACTTTCAGATCAAGATCCAGTACAGCGGC 1407

Db 623 GTGCTGATGGAGACAGACAGCGCCCTTCCACTTCAAGATCAAGATCCCGAAGTAAGCGC 682

Qy 1408 TACGAGGTGCCCTTGAGAGACCCCGGTGTGCCACAGCGCGGACCCGCTCCCACTTACAGC 1467

Db 683 TACGAGGTGCCCTTGAGAGACCCCAAGTGTGTGAGCAGGACCACTCCCACTTACAGC 742

Qy 1468 GTGGAGTTCTCCGAGAGCCCTTCGCGGTGTGATGTGTGCAACCGGAGTGGAGCGCCCGTG 1527

Db 743 GTGGAATTCTCAGAGGAACCTTTGGAGTGTGCTGTGAGGAAGCTTGTGTGCGCGAGTG 802

Qy 1528 CTGCTGAACACAGCGGTGGCGCCCTGTCTTTTGGGACAGTTCCTTTCAGTGTCCACC 1587

Db 803 TTGCTGAACACAAACCGTGGCCCGCTGTCTTTCGTGACCAAGTTCCTGAGTGTCCACT 862

Qy 1588 TCGCTGCCCTCGCAGTATATCACAGCCCTCGCGGACCTCAGTCCCTGTGTGCTCAGC 1647

Db 863 TCCCTGCCCGCCAGCAGCATCAAGCCCTGGGGGAACACTCAGCCCACTCATGTCTCAGC 922

DT	18-NOV-2004	(first entry)	
XX	Human diagnostic and therapeutic polynucleotide SEQ ID NO:1844.		
DE	ss; gene; gene therapy; human diagnostic and therapeutic polynucleotide;		
KW	dithp.		
XX	Homo sapiens.		
OS	WO2004023973-A2.		
XX	25-MAR-2004.		
PN	12-SEP-2003; 2003WO-US028227.		
XX	12-SEP-2002; 2002US-0410259P.		
PF	12-SEP-2002; 2002US-0410260P.		
PR	(INCY-) INCYTE CORP.		
XX	Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F, Harthshorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV; Mooney EM, Deleagane AM, Panesar IS, Banville SC, Reddy RP; Stevens KA, Blanchard JL, Panux SR, Wang X, Au AP, Gerstein EH; Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL; Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirtan ES; Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D; Patury S, Shi X, Suarez CJ;		
PI	WPI; 2004-329368/30.		
XX	P-PSDB; ABM84317.		
DR	New diagnostic and therapeutic polynucleotides and polypeptides, useful in diagnosing a condition, disease or disorder associated with human molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or in gene mapping.		
PT	Claim 1; Page; 190pp; English.		
XX	The invention relates to novel diagnostic and therapeutic polynucleotides selected from one of the 2722 sequences defined in the specification. A polynucleotide of the invention may have a use in gene therapy. The human diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be used to diagnose a particular condition, disease or disorder associated with human molecules, e.g. cell proliferative disorders, autoimmune/inflammatory disorder, developmental disorder, endocrine disorder, neurological disorders, gastrointestinal disorders, or infections caused by virus, bacteria, fungi or parasite. The dithp molecules may also be used in genetic mapping, in identifying individuals from minute biological samples, in detecting single nucleotide polymorphisms, as molecular weight markers, and for somatic or germline gene therapy. The present sequence represents a dithp polynucleotide of the invention. Note: The sequence data for this patent is not represented in the printed specification, but was obtained in electronic format directly from WIPO at www.wipo.int/pct/en/sequences/listing.htm		
CC	Sequence 2655 BP; 455 A; 900 C; 814 G; 486 T; 0 U; 0 Other;		
SQ	Query Match 32.2%; Score 1191.6; DB 13; Length 2655; Best Local Similarity 93.6%; Pred. No. 7.3e-214; Matches 1286; Conservative 0; Mismatches 9; Indels 79; Gaps 1;		
QY	1036	GAACCGGTGCACACCCCGCGTCCAGAGCAGTGCCACACAGTGGCAGCTCCCGCC 1095	
DB	454	GATGCCAGGACACCCCGCGCTCCAGAGCAGTGCCACACAGTGGCAGCTCCCGCC 513	
QY	1096	AACAGCGGTTCGATTCGCGCCCTGCAAGGCCATCACCCAGGAAACAGTGCAGGCGCCG 1155	
DB	514	AACAGCGGTTCGATTCGCGCCCTGACAGGCCATCACCCAGGAAACAGTGCAGGCGCCG 573	
QY	1156	GGCTGTGTACATCCCTGCAAGCAAGGCGGTGTCAGGAGGCCAGATGGGGCAGCCCTGG 1215	
DB	574	GGCTGTGTACATCCCTGCAAGCAAGGCGGTGTCAGGAGGCCAGATGGGGCAGCCCTGG 1333	
QY	1216	TGCTTTTCCCAACCAGCTACCCCAAGCTACAAGCTGGAGAACTGAGTCTCTCTGAAATG 1275	
DB	634	TGCTTTTCCCAACCAGCTACCCCAAGCTACAAGCTGGAGAACTGAGTCTCTCTGAAATG 693	
QY	1276	GGCTACAGGGCCACCTGACCCCGTACCAACCCCAAGCTTTTCCCAAGGACATCTGACC 1335	
DB	694	GGCTACAGGGCCACCTGACCCCGTACCAACCCCAAGCTTTTCCCAAGGACATCTGACC 753	
QY	1336	CTGCGGCTGGAGCTGATGATGAGAGCTGAGAAACCGCTCCACTTACAGATCAAGATCCA 1395	
DB	754	CTGCGGCTGGAGCTGATGATGAGAGCTGAGAAACCGCTCCACTTACAGATCAAGATCCA 813	
QY	1396	GCTAACAGCGCTACGAGGTGCCCTTGGAGACCCCGCGTGTCCACAGCGGCGACCGTCC 1455	
DB	814	GCTAACAGCGCTACGAGGTGCCCTTGGAGACCCCGCATGTCCACAGCGGCGACCGTCC 873	
QY	1456	CCACTCTACAGCGTGGAGTTCTCCGAGGAGCCCTTGGGGGTGATCGTGGCGGCGAGCTG 1515	
DB	874	CCACTCTACAGCGTGGAGTTCTCCGAGGAGCCCTTGGGGGTGATCGTGGCGGCGAGCTG 933	
QY	1516	GACGCGCGGTCTGCTGAACACGAGCGGTGGCGCCCTGTTCTTTCGCGACAGTTCCCTT 1575	
DB	934	GACGCGCGGTCTGCTGAACACGAGCGGTGGCGCCCTGTTCTTTCGCGACAGTTCCCTT 993	
QY	1576	CAGCTGTCCACCTCGCTGCCCTCGCAGTATATCACAGGCTTCGCGAGCACCTCAGTCCC 1635	
DB	994	CAGCTGTCCACCTCGCTGCCCTCGCAGTATATCACAGGCTTCGCGAGCACCTCAGTCCC 1053	
QY	1636	CTGATGCTCAGCACAGCTGGACAGGATCACCCCTGTGTGAACCGGACCTTGGCGCCACG 1695	
DB	1054	CTGATGCTCAGCACAGCTGGACAGGATCACCCCTGTGTGAACCGGACCTTGGCGCCACG 1113	
QY	1696	CCCGGTGGGAACTCTACGGGTCTCACCCCTTTTACCTGGCGCTGGAGACGCGGGTGC 1755	
DB	1114	CCCGGTGGGAACTCTACGGGTCTCACCCCTTTTACCTGGCGCTGGAGACGCGGGTGC 1173	
QY	1756	GCACACGGGGTCTCTGCTAAACAGCAATGCATGGATGTGGTCTCTGAGCGGAGCCCT 1815	
DB	1174	GCACACGGGGTCTCTGCTAAACAGCAATGCATGGATGTGGTCTCTGAGCGGAGCCCT 1233	
QY	1816	GCCCTTAGCTGAGGTGACAGGTGGATCTCTGGATGTACATCTTCTGGGCGCCAGAG 1875	
DB	1234	GCCCTTAGCTGAGGTGACAGGTGGATCTCTGGATGTACATCTTCTGGGCGCCAGAG 1293	
QY	1876	CCCAAGAGCGTGTGTCAGCAGTACCTGGACGTGTG----- 1911	
DB	1294	CCCAAGAGCGTGTGTCAGCAGTACCTGGACGTGTG----- 1353	
QY	1912	-----GGATA 1916	
DB	1354	GCCCCCGCCCCAAGGCTCCCTCCCTCCCTCATGAAGTCGGCGTGGCTGCAGGATA 1413	
QY	1917	CCCGTTATGCGCGCATACTGGGGCTGGGCTTCCACTGTGCGCTGGGGCTACTCTC 1976	
DB	1414	CCCGTTATGCGCGCATACTGGGGCTGGGCTTCCACTGTGCGCTGGGGCTACTCTC 1473	
QY	1977	CACCGCTATCACCCCGCGAGTGTGGAGAAATGACACAGGCGCCACTTCCCTCGACGT 2036	
DB	1474	CACCGCTATCACCCCGCGAGTGTGGAGAAATGACACAGGCGCCACTTCCCTCGACGT 1533	
QY	2037	CCAAATGAACGACCTGGAGTACATGGACTCCCGGAGGAGCTTTCACGTTTCAACAAGGATG 2096	
DB	1534	CCAGTGAACGACCTGGAGTACATGGACTCCCGGAGGAGCTTTCACGTTTCAACAAGGATG 1593	
QY	2097	CTTCCGGGACTTCCCGGCGATGGTGCAGAGTGCACAGGGGCGCGCGCTACATGAT 2156	
DB	1594	CTTCCGGGACTTCCCGGCGATGGTGCAGAGTGCACAGGGGCGCGCGCTACATGAT 1653	
QY	2157	GATCGTGGATCTGCGCATCAGCAGCTCGCGCCCTGCGGGAGCTACAGGCCCTACGACCA 2216	
DB	1654	GATCGTGGATCTGCGCATCAGCAGCTCGCGCCCTGCGGGAGCTACAGGCCCTACGACCA 1713	

QY 2217 GGGTCTGGAGGGGGTGTTCATCAACAAAGAGACCGCCAGCCGCTGATTGGGAAGGT 2276
DB |||||
1714 GGGTCTGGAGGGGGTGTTCATCAACAAAGAGACCGCCAGCCGCTGATTGGGAAGGT 1773
QY 2277 ATGGCCCGGGTCCACTGCTTCCCGCACTTCCACCAACCCACAGCCCTGGCCTG 2330
DB |||||
1774 ATGGCCCGGGTCCACTGCTTCCCGCACTTCCACCAACCCACAGCCCTGGCCTG 1827

RESULT 11

AAC87527

ID AAC87527 standard; DNA; 1493 BP.

XX

AC

XX

AC

XX

DT

XX

DE

XX

KW

KW

KW

KW

XX

OS

XX

XX

PN

XX

PD

XX

PF

XX

PR

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Human RAP (receptor associated protein) cDNA.

RAP; receptor associated protein; human; factor VIII clearance;

LRP-mediated plasma clearance; receptor-dependent clearance;

receptor-independent clearance; ligand internalisation;

low density lipoprotein related protein; haemophilia; half-life; ss.

Homo sapiens.

WO200071714-A2.

30-NOV-2000.

24-MAY-2000; 2000WO-US014111.

24-MAY-1999; 99US-0135847P.

(AMNA-) AMERICAN NAT RED CROSS.

Saenko EL, Strickland DK;

WPI; 2001-025163/03.

P-PSDB; AAB48844.

Factor VIII mutants having increased half-life useful for treating

hemophilia, comprise one or more amino acid substitutions in the A2

and/or C2 domain of factor VIII.

Disclosure; Page 111-114; 121pp; English.

The invention relates to human factor VIII mutants comprising an amino

acid substitution at one or more positions in the A2 domain and/or an

amino acid substitution at one or more positions in the C2 domain. The

invention also encompasses a factor VIII mutant which lacks a B domain

(AAB48842). The factor VIII mutants have an increased half-life in the

bloodstream. The A2 domain mutants exhibit reduced LRP-dependent

(receptor-dependent) clearance of factor VIII, while C2 domain mutants

have reduced receptor-independent clearance. The invention also relates

to a method of using RAP (receptor associated protein), a protein which

inhibits LRP (low density lipoprotein related protein)-mediated ligand

internalisation, to increase the half-life of factor VIII. The mutant

factor VIII proteins, and nucleotides encoding them, are useful for

treating haemophilia. RAP, LRP-binding RAP mutants or fragments, and

nucleic acids encoding them may also be used in the treatment of

haemophilia, in combination with a mutant factor VIII protein or DNA of

the invention. The invention provides means of increasing the half-life

of factor VIII by reducing its clearance from plasma. The present

sequence represents cDNA encoding human RAP

Sequence 1493 BP; 360 A; 412 C; 488 G; 233 T; 0 U; 0 Other;

Query Match 25.9%; Score 957; DB 5; Length 1493;

Best Local Similarity 99.0%; Pred. No. 6.7e-170;

Matches 963; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 63 CGGATCCTACTCGCGGAGAGAAACCGAGCCCGTCCCGGAAACGGCGAGTCCGGAGA 122

DB |||||
109 CGGCAAGTACTCGCGGAGAAAGAACAGCCCAAGCCGTCGCCAAACGCGAGTCCGGAGA 168
QY |||||
123 GGAGTTCCGCACTGGAGAAAGTTGAACACAGCTGTGGGAGAAAGGCCAGCGACTGCATCTTCC 182
DB |||||
169 GGAGTTCCGCACTGGAGAAAGTTGAACACAGCTGTGGGAGAAAGGCCAGCGACTGCATCTTCC 228
QY |||||
183 TCCCGTGGAGGCTGGCCGAGCTCCACGCTGATCTGAAGATACAGAGAGAGGAGCGAACTCGC 242
DB |||||
229 TCCCGTGGAGGCTGGCCGAGCTCCACGCTGATCTGAAGATACAGAGAGAGGAGCGAACTCGC 288
QY |||||
243 CTGGAAGAAACTAAAGCTTTGACGGCTTGGACGAAAGATGGGAGAAAGAGAGCGAGACTCAT 302
DB |||||
289 CTGGAAGAAACTAAAGCTTTGACGGCTTGGACGAAAGATGGGAGAAAGAGAGCGAGACTCAT 348
QY |||||
303 ACGCAACCTCAATGTCTTGGCCAAAGTATGCTGTGGACGGAAGAGAGAGCGCTCGGCA 362
DB |||||
349 ACGCAACCTCAATGTCTTGGCCAAAGTATGCTGTGGACGGAAGAGAGAGCGCTCGGCA 408
QY |||||
363 GGTGACCAAGCAACTCCCTCAGTGGCACCAGGAAGACGGGCTGGATGACCCAGGCTGGA 422
DB |||||
409 GGTGACCAAGCAACTCCCTCAGTGGCACCAGGAAGACGGGCTGGATGACCCAGGCTGGA 468
QY |||||
423 AAAGCTGTGGCAACAAGGCGAAGACCTCTGGGAAATTTCTCCGCGAAGAACTGGACAAGCT 482
DB |||||
469 AAAGCTGTGGCAACAAGGCGAAGACCTCTGGGAAATTTCTCCGCGAAGAACTGGACAAGCT 528
QY |||||
483 CTGGCGGAGTTCCTGTCATCAAAAGAGAAAGTTCCAGAGTACCAACGCTCTCTCTGGAGAC 542
DB |||||
529 CTGGCGGAGTTCCTGTCATCAAAAGAGAAAGTTCCAGAGTACCAACGCTCTCTCTGGAGAC 588
QY |||||
543 CCTGAGCAGGACCGGAAGAAATTCACGAGAAAGCTCTAGCCCTCGGACCTCGAGCGACAT 602
DB |||||
589 CCTGAGCAGGACCGGAAGAAATTCACGAGAAAGCTCTAGCCCTCGGACCTCGAGCGACAT 648
QY |||||
603 CAAAGGCGAGCTCTGTCACAGCAGCACAACGAGCTGAAGAGAGAGCTGGCGAGCATCAA 662
DB |||||
649 CAAAGGCGAGCTCTGTCACAGCAGCACAACGAGCTGAAGAGAGAGCTGGCGAGCATCAA 708
QY |||||
663 CCAGGCGCTTGGACCGCCCTGCGCAGGCTCAGCCACAGGCTACAGCACTGAGGCTGAGTT 722
DB |||||
709 CCAGGCGCTTGGACCGCCCTGCGCAGGCTCAGCCACAGGCTACAGCACTGAGGCTGAGTT 768
QY |||||
723 CGAGGAGCCAGGCTGATTTGACCTGTGGACCTGGCGAGTCCGCCAACCTCACGGACAA 782
DB |||||
769 CGAGGAGCCAGGCTGATTTGACCTGTGGACCTGGCGAGTCCGCCAACCTCACGGACAA 828
QY |||||
783 GGAGCTGGAGGCTTCCGCGAGGAGCTCAGCACTTCGAGCCAAATCGAAGACGACAA 842
DB |||||
829 GGAGCTGGAGGCTTCCGCGAGGAGCTCAGCACTTCGAGCCAAATCGAAGACGACAA 888
QY |||||
843 CCAGTACCAAGAGCAGCTGGAGATTGCGCAGAGAAAGCTGAGGCAACGAGAGAGCGTGGG 902
DB |||||
889 CCAGTACCAAGAGCAGCTGGAGATTGCGCAGAGAAAGCTGAGGCAACGAGAGAGCGTGGG 948
QY |||||
903 CGACCGGAGCGTGTGAGCCGCGAGAGAAAGCAACGCTGTGGAGGGGCGGACCAA 962
DB |||||
949 CGACCGGAGCGTGTGAGCCGCGAGAGAAAGCAACGCTGTGGAGGGGCGGACCAA 1008
QY |||||
963 GGAGCTGGGCTACAGGCTGAAGAGCACTTCGAGCACTGTCCGCGAGGATCTCCAGAGC 1022
DB |||||
1009 GGAGCTGGGCTACAGGCTGAAGAGCACTTCGAGCACTGTCCGCGAGGATCTCCAGAGC 1068
QY |||||
1023 TCGCCCGGAGGCA 1035
DB |||||
1069 TCGGCACAAACGAA 1081

RESULT 12

AAL48893

ID AAL48893 standard; cDNA; 1493 BP.

XX

AC AAL48893;

XX	24-OCT-2002 (first entry)	
DT	Human receptor-associated protein coding sequence.	
XX	Human; factor VIII; fVIII; half-life; mutant; haemophilia;	
DE	heparan sulfate proteoglycan-mediated clearance; RAP;	
XX	receptor-associated protein; gene; ss; haemostatic; gene therapy;	
KW	alpha2 macroglobulin receptor-associated protein.	
KW		
XX	Homo sapiens.	
OS		
XX		
PH	Key Location/Qualifiers	
FT	CDS 14..1087	
FT	/*tag= a	
FT	/product= "RAP"	
FT	sig_peptide 14..115	
FT	/*tag= b	
FT	mat_peptide 116..1084	
FT	/*tag= c	
XX	WO200260951-A2.	
PN		
XX	08-AUG-2002.	
PD		
XX	11-JAN-2002; 2002WO-US000583.	
PF		
XX	12-JAN-2001; 2001US-0260904P.	
PR		
XX	(AMNA-) AMERICAN NAT RED CROSS.	
PA		
XX	Saenko EL, Sarafanov AG;	
PI		
XX	WPI; 2002-608501/65.	
DR	P-PSDB; AA018621.	
DR		
XX	New mutant factor VIII with reduced sulfate proteoglycan (HSPG)-dependent	
PT	or receptor-independent clearance and procoagulant activity for treating	
PT	hemophilia.	
PT		
XX	Disclosure; Page 151-154; 161pp; English.	
PS		
XX	The present invention relates to a mutant factor VIII protein with	
CC	reduced sulfate proteoglycan (HSPG)-dependent or receptor-independent	
CC	clearance and procoagulant activity, which has a nonconservative amino	
CC	acid substitution at one or more positions in the A2 domain consisting of	
CC	Lys(380, 512, 556, 570 or 659) or Arg(490, 527, 562 or 571) or in the C2	
CC	domain relative to the wild-type. The mutant factor VIII or the	
CC	polynucleotide encoding it and a receptor-associated protein (alpha2	
CC	macroglobulin receptor-associated protein or RAP) are useful for treating	
CC	haemophilia. The mutated protein has a longer half-life. The present	
CC	sequence is the human receptor-associated protein coding sequence	
XX		
SQ	Sequence 1493 BP; 360 A; 412 C; 488 G; 233 T; 0 U; 0 Other;	
Query Match 25.9%; Score 957; DB 6; Length 1493;		
Best Local Similarity 99.0%; Pred. No. 6.7e-170;		
Matches 963; Conservative 0; Mismatches 10; Indels 0; Gaps 0;		
QY	63 CGGATCCTACTCGCGGAGAGAACAGCCGCAAGCGCTCCCGAAACGGAGTCCGGAGA 122	
DB	109 CGGCAAGTACTCGCGGAGAGAACAGCCGCAAGCGCTCCCGAAACGGAGTCCGGAGA 168	
QY	123 GGAGTTCGCGATGAGAGTTGAACAGCTGTGGAGAGGCCCGCGACTGCTCTTC 182	
DB	169 GGAGTTCGCGATGAGAGTTGAACAGCTGTGGAGAGGCCCGCGACTGCTCTTC 228	
QY	183 TCCCGTGAGGCTCGCGAGCTCCACGCTGATCTGAAGATACAGGAGGAGCGAACTCGC 242	
DB	229 TCCCGTGAGGCTCGCGAGCTCCACGCTGATCTGAAGATACAGGAGGAGCGAACTCGC 288	
QY	243 CTGGAAGAACTAAAGCTTGACCGCTTGACGGAAGATGGGGAGAGGAGCGAGACTCAT 302	

Db	289	CTGGAGAAACTAAAGCTTGACGGCTTGACGAGATGGGAGAGGAAGCGAGACTCAT	348
QY	303	ACGCAACTCAATGTCACTTGGCCCAAGTATGGTCTGGACGGAAGAAAGACGCTCGGCA	362
Db	349	ACGCAACTCAATGTCACTTGGCCCAAGTATGGTCTGGACGGAAGAAAGACGCTCGGCA	408
QY	363	GGTGACCAAGCACTCCCTCAGTGGGACCCAGGAGAGCGGGCTGGATGACCCAGGCTGGA	422
Db	409	GGTGACCAAGCACTCCCTCAGTGGGACCCAGGAGAGCGGGCTGGATGACCCAGGCTGGA	468
QY	423	AAAGCTGTGGCAAGGCGAAGACCTCTGGGAAATTTCTCCGCGAAGAACTTGACAAAGCT	482
Db	469	AAAGCTGTGGCAAGGCGAAGACCTCTGGGAAATTTCTCCGCGAAGAACTTGACAAAGCT	528
QY	483	CTGGCGGAGTTCCTGTCATCAAAAGAGAAAGTTTCAACGAGTCAACGCTCTGCTGGAGAC	542
Db	529	CTGGCGGAGTTCCTGTCATCAAAAGAGAAAGTTTCAACGAGTCAACGCTCTGCTGGAGAC	588
QY	543	CCTGAGCAGGACCGAAGAAATCCAGGAACTTATTAGCCCTCGGACCTGAGCGACAT	602
Db	589	CCTGAGCAGGACCGAAGAAATCCAGGAACTTATTAGCCCTCGGACCTGAGCGACAT	648
QY	603	CAAGGCGAGCTCTGTCACAGCAGGACACCGAGCTGAAGAGAGAACTCGCGACATCAA	662
Db	649	CAAGGCGAGCTCTGTCACAGCAGGACACCGAGCTGAAGAGAGAACTCGCGACATCAA	708
QY	663	CCAGGCGCTTGGACCGCTTGGCAGGCTCAGCCACAGGCTTACAGCACTGAGGCTGAGTT	722
Db	709	CCAGGCGCTTGGACCGCTTGGCAGGCTCAGCCACAGGCTTACAGCACTGAGGCTGAGTT	768
QY	723	CGAGAGCCCGAGGTGATTGACCTGTGGGACCTGGCGGAGTCCGCCAATCTCACGACAA	782
Db	769	CGAGAGCCCGAGGTGATTGACCTGTGGGACCTGGCGGAGTCCGCCAATCTCACGACAA	838
QY	783	CGAGCTTGGAGGCGTTCCCGGAGGAGCTCAAGCACTTTCGAAGCCAAATCGAGAAGCACA	842
Db	829	GGAGCTTGGAGGCGTTCCCGGAGGAGCTCAAGCACTTTCGAAGCCAAATCGAGAAGCACA	888
QY	843	CCACTACCAAGAGCAGCTGGAGATTGCGCAACGAGAGCTGAGGACCGCAGAGAGCGTGGG	902
Db	889	CCACTACCAAGAGCAGCTGGAGATTGCGCAACGAGAGCTGAGGACCGCAGAGAGCGTGGG	948
QY	903	CGACGCGAGCTGTGAGCCGCGAGAGCAAGCAAGCCCTGCTGGAGGCGCGGACCAA	962
Db	949	CGACGCGAGCTGTGAGCCGCGAGAGCAAGCAAGCCCTGCTGGAGGCGCGGACCAA	1008
QY	963	CGAGCTGGGCTACACGCTGAAGCAATCTGAGAGACCTGTCGGCAGGATCTCCAGAGC	1022
Db	1009	GGAGCTGGGCTACACGCTGAAGCAATCTGAGAGACCTGTCGGCAGGATCTCCAGAGC	1068
QY	1023	TCGCGCCGAGGCA 1035	
Db	1069	TCGCGCACACGAA 1081	

RESULT 13
ABK84062
ID ABK84062 standard; cDNA; 1493 BP.
XX
AC ABK84062;
XX
DT 14-AUG-2002 (first entry)
XX
DE Human cDNA differentially expressed in granulocytic cells #633.
XX
KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW viral infection; parasitic infection; protozoal infection;
KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.

XX 05-AUG-2004.
XX 15-JAN-2004; 2004WO-US001100.
XX 15-JAN-2003; 2003US-00342887.
XX (ROSE-) ROSETTA INPHARMATICS LLC.
XX (NECA-) NETHERLANDS CANCER INST.
XX Van't Veer LJ, He Y;
XX WPI; 2004-593473/57.
XX
XX Classifying a breast cancer patient according to prognosis comprises
XX determining the similarity between the level of expression of each of
XX five genes in a cell sample taken from patient, to control levels.
XX
XX Disclosure; SEQ ID NO 757; 226pp; English.
XX
XX The invention relates to a method of classifying a breast cancer patient
XX according to prognosis by determining the similarity between the level of
XX expression of each of five genes for which markers are listed in the
XX specification, in a cell sample taken from the breast cancer patient, to
XX control levels of expression for each respective five genes to obtain a
XX patient similarity value. The methods are useful for classifying a breast
XX cancer patient according to prognosis. Kits and computer program products
XX are useful for data analysis using the diagnostic, prognostic and
XX statistical methods of the invention. This sequence corresponds to a
XX marker used in the method of the invention.
XX
XX Sequence 1493 BP; 360 A; 412 C; 488 G; 233 T; 0 U; 0 Other;
XX
Query Match 25.9%; Score 957; DB 13; Length 1493;
Best Local Similarity 99.0%; Pred. No. 6.7e-170;
Matches 963; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
XX 63 CGGATCCTACTCGCGGAGAGAAACAGCCAGCCGTCCTCCGMAACGGAGTCCGGAGA 122
XX 109 CGGCAAGTACTCGCGGAGAGAGAACAGCCAGCCGTCCTCCGMAACGGAGTCCGGAGA 168
XX 123 GGAGTTCGCGATGGAGAAAGTTGAACAGAGTGTGGAGAAAGCCAGCGACTGCATCTTCC 182
XX 169 GGAGTTCGCGATGGAGAAAGTTGAACAGAGTGTGGAGAAAGCCAGCGACTGCATCTTCC 228
XX 183 TCCGCTGAGGCTGGCGGAGCTCCAGCTGATCTGAGATACAGGAGAGGAGCGAACTCGC 242
XX 229 TCCCGTGAGGCTGGCGGAGCTCCAGCTGATCTGAGATACAGGAGAGGAGCGAACTCGC 288
XX 243 CTGGAGAGAACTAAAGCTTGAGCGGCTTGAGCGAAGATGGGAGAGAGGAGCGAGACTCAT 302
XX 289 CTGGAGAGAACTAAAGCTTGAGCGGCTTGAGCGAAGATGGGAGAGAGGAGCGAGACTCAT 348
XX 303 AGCGAACCTCAATGTGTCATTTGGCCAAAGTATGGTCTGGACGGAAGAGGAGCGCTCGGCA 362
XX 349 AGCGAACCTCAATGTGTCATTTGGCCAAAGTATGGTCTGGACGGAAGAGGAGCGCTCGGCA 408
XX 363 GGTGACCAAGCACTCCCTCAGTGGGACCCAGGAGAGCGGCTGGATGACCCCGAGGCTGGA 422
XX 409 GGTGACCAAGCACTCCCTCAGTGGGACCCAGGAGAGCGGCTGGATGACCCCGAGGCTGGA 468
XX 423 AAAGCTGTGGCACAAGCGCAAGCACTCTGGGAAATTTCTCCGCGAAGAACTGGACAAGCT 482
XX 469 AAAGCTGTGGCACAAGCGCAAGCACTCTGGGAAATTTCTCCGCGAAGAACTGGACAAGCT 528
XX 483 CTGGCGGGAGTTCTCTGCATCAAAAGAGAAAGTTACAGATCAACGTCCTCTGTGGAGAC 542
XX 529 CTGGCGGGAGTTCTCTGCATCAAAAGAGAAAGTTACAGATCAACGTCCTCTGTGGAGAC 588
XX 543 CTTGAGCAGGACCGAGAAATCCACGAGAAAGCTCATTAGCCCTCGGACCTGAGCGACAT 602
XX 589 CTTGAGCAGGACCGAGAAATCCACGAGAAAGCTCATTAGCCCTCGGACCTGAGCGACAT 648

QY 603 CAAGGGCAGCGTCTTGCA CAGCAGGCACACGAGAGCTGAAGGAGAAAGCTGCGCAGCATCAA 662
DB 649 CAAGGGCAGCGTCTTGCA CAGCAGGCACACGAGAGCTGAAGGAGAAAGCTGCGCAGCATCAA 708
QY 663 CCAGGGCCTTGACCCCTGCGCAGGAGGTCAAGCCACGAGGCTACAGCACTGAGGCTGAGTT 722
DB 709 CCAGGGCCTTGACCCCTGCGCAGGAGGTCAAGCCACGAGGCTACAGCACTGAGGCTGAGTT 768
QY 723 CGAGGAGCCCAAGGCTGATTGACCTGCTGGGACCTGCGCAGTCCGCAACCTCAACGACAA 782
DB 769 CGAGGAGCCCAAGGCTGATTGACCTGCTGGGACCTGCGCAGTCCGCAACCTCAACGACAA 828
QY 783 GGAGCTGGAGGCGTTTCCGGGAGGAGCTCAAGCACTTCGAAGCCAAATCGAAGACACAA 842
DB 829 GGAGCTGGAGGCGTTTCCGGGAGGAGCTCAAGCACTTCGAAGCCAAATCGAAGACACAA 888
QY 843 CCAGTACCAAGAGCAGCTGGAGATTGCCACGAGAGCTGAGGACCGCAGAGAGCGTGGG 902
DB 889 CCAGTACCAAGAGCAGCTGGAGATTGCCACGAGAGCTGAGGACCGCAGAGAGCGTGGG 948
QY 903 CGAGCGCGAGCGTGTGAGCCGCGAGAGCAACGCCCTGCTGGAGGGGCGGACCAA 962
DB 949 CGAGCGCGAGCGTGTGAGCCGCGAGAGCAACGCCCTGCTGGAGGGGCGGACCAA 1008
QY 963 GGAGCTGGGCTACACGCTGAAGAGCACTTCGAGGACCTGTCCGGCAGGATCTCCAGAGC 1022
DB 1009 GGAGCTGGGCTACACGCTGAAGAGCACTTCGAGGACCTGTCCGGCAGGATCTCCAGAGC 1068
QY 1023 TCGCGCCGAGGCA 1035
DB 1069 TCGGCACAACGAA 1081

Search completed: April 4, 2005, 11:37:46
JOB time : 1774 secs

Result No.	Score	Query Match	Length	DB ID	Description	
1	2647.8	71.5	3846	4	US-09-949-016-23	Sequence 23, Appl
2	2644.6	71.4	3847	4	US-09-949-016-3656	Sequence 3656, Ap
3	370.6	10.0	449	3	US-09-997-787-175	Sequence 175, App
4	369.8	10.0	2752	1	US-08-030-925A-3	Sequence 3, Appli
5	340.4	9.2	22404	4	US-09-949-016-11765	Sequence 11765, A
6	340.4	9.2	22404	4	US-09-949-016-15398	Sequence 15398, A
7	212.2	5.7	2272	4	US-09-744-926-1	Sequence 1, Appli
8	189.8	5.1	3018	4	US-09-949-016-4756	Sequence 4756, Ap
9	175.2	4.7	601	4	US-09-949-016-18094	Sequence 18094, A
10	175.2	4.7	601	4	US-09-949-016-128902	Sequence 128902, A
11	165.2	4.5	601	4	US-09-949-016-128901	Sequence 18093, A
12	165.2	4.5	601	4	US-09-949-016-128901	Sequence 128901, A
13	157.4	4.3	601	4	US-09-949-016-18081	Sequence 18081, A
14	157.4	4.3	601	4	US-09-949-016-128889	Sequence 128889, A
15	155	4.2	601	4	US-09-949-016-18077	Sequence 18077, A
16	155	4.2	601	4	US-09-949-016-128885	Sequence 128885, A
17	153.8	4.2	601	4	US-09-949-016-18075	Sequence 18075, A
18	153.8	4.2	601	4	US-09-949-016-18076	Sequence 18076, A
19	153.8	4.2	601	4	US-09-949-016-128883	Sequence 128883, A
20	153.8	4.2	601	4	US-09-949-016-128884	Sequence 128884, A
21	153.4	4.1	601	4	US-09-949-016-18078	Sequence 18078, A
22	153.4	4.1	601	4	US-09-949-016-18079	Sequence 18079, A
23	153.4	4.1	601	4	US-09-949-016-128886	Sequence 128886, A
24	153.4	4.1	601	4	US-09-949-016-128887	Sequence 128887, A
25	149.6	4.0	601	4	US-09-949-016-18080	Sequence 18080, A
26	149.6	4.0	601	4	US-09-949-016-128888	Sequence 128888, A
27	135.8	3.7	601	4	US-09-949-016-18082	Sequence 18082, A

Db 940 |||||CTGCGCTGAGACGTGATGGAGACTGAGAACCGCTCCATCTTACGATCAAGATCCA 999
QY 1396 GCTTAACAGCGCTACAGAGTCCCTTGGAGACCCCGGTGTCCACAGCGCGGACCGTCC 1455
Db 1000 GCTTAACAGCGCTACAGAGTCCCTTGGAGACCCCGGTGTCCACAGCGCGGACCGTCC 1059
QY 1456 CCACTCTACAGCGTGGAGTTCTCCGAGGAGCCCTTCGGGGTGATCGTGACCGGACAGCTG 1515
Db 1060 CCACTCTACAGCGTGGAGTTCTCCGAGGAGCCCTTCGGGGTGATCGTGACCGGACAGCTG 1119
QY 1516 GACGGCGCGTGTCTGTAACAGAGCGGTGGCGCCCTGTGTTCTTTCGGGACGAGTTCCCTT 1575
Db 1120 GACGGCGCGTGTCTGTAACAGAGCGGTGGCGCCCTGTGTTCTTTCGGGACGAGTTCCCTT 1179
QY 1576 CAGCTGTCCACCTCGCTGCCCTCGCAGTATATCACAGGCGCTCGCCAGCACCTCAGTCCC 1635
Db 1180 CAGCTGTCCACCTCGCTGCCCTCGCAGTATATCACAGGCGCTCGCCAGCACCTCAGTCCC 1239
QY 1636 CTGATGCTCAGCACACAGCTGGACAGGATCACCTGTGGAACCGGACCTTGGCCCAAG 1695
Db 1240 CTGATGCTCAGCACACAGCTGGACAGGATCACCTGTGGAACCGGACCTTGGCCCAAG 1299
QY 1696 CCGGTGCGAACTCTACGGGTCTACCTTTCTACTCTGGCGTGAAGAGCGCGGGTCG 1755
Db 1300 CCGGTGCGAACTCTACGGGTCTACCTTTCTACTCTGGCGTGAAGAGCGCGGGTCG 1359
QY 1756 GCACAGGGGTGTTCTGTCTAAACAGCAATGCCATGGATGTGTCTGCAGCGGAGCCCT 1815
Db 1360 GCACAGGGGTGTTCTGTCTAAACAGCAATGCCATGGATGTGTCTGCAGCGGAGCCCT 1419
QY 1816 GCCTTTAGCTGGAGGTGACAGGTGGATCTCGATGTCTACATCTTCTGGGCCACAGAG 1875
Db 1420 GCCCTTAGCTGGAGGTGACAGGTGGATCTCGATGTCTACATCTTCTGGGCCACAGAG 1479
QY 1876 CCACAGAGCGTGTGACAGATGACTGAGCTGTGGATGATACCGGTTATGCGCGCATAC 1935
Db 1480 CCACAGAGCGTGTGACAGATGACTGAGCTGTGGATGATACCGGTTATGCGCGCATAC 1539
QY 1936 TGGGGCTTGGCTTCCACCTGTGCGCTGGGGTACTCTCCACCGCTATACCGGCCAG 1995
Db 1540 TGGGGCTTGGCTTCCACCTGTGCGCTGGGGTACTCTCTCCACCGCTATACCGGCCAG 1599
QY 1996 GTGGTGGAGAACATGACAGGGGCCCTTCCCGCTGGAGCTCCAAATGGAACGACCTGGAC 2055
Db 1600 GTGGTGGAGAACATGACAGGGGCCCTTCCCGCTGGAGCTCCAAATGGAACGACCTGGAC 1659
QY 2056 TACATGGACTCCGGAGGAGCTTCAAGTTCAACAGGATGGCTTCCGGGACTTCCCGGCC 2115
Db 1660 TACATGGACTCCGGAGGAGCTTCAAGTTCAACAGGATGGCTTCCCGGACTTCCCGGCC 1719
QY 2116 ATGGTCAGGAGCTGCACAGGGCGCGCGCTACATGATGATCGTGGATCCTGCCATC 2175
Db 1720 ATGGTCAGGAGCTGCACAGGGCGCGCGCTACATGATGATCGTGGATCCTGCCATC 1779
QY 2176 AGCAGCTCGGGCCCTCGCGGAGCTACAGCCCTACGACAGGGTCTCGGAGGGGGTT 2235
Db 1780 AGCAGCTCGGGCCCTCGCGGAGCTACAGCCCTACGACAGGGTCTCGGAGGGGGTT 1839
QY 2236 TTCAATCAACAGACCGGCCAGCGCTGATTTGGGAAGGATATGGCCCGGGTCCAATGCC 2295
Db 1840 TTCAATCAACAGACCGGCCAGCGCTGATTTGGGAAGGATATGGCCCGGGTCCAATGCC 1899
QY 2296 TTCCCGGACTTCAACAGCCCAAGCCCTGGCTGTGGGAGACATGGTGGTGAATTC 2355
Db 1900 TTCCCGGACTTCAACAGCCCAAGCCCTGGCTGTGGGAGACATGGTGGTGAATTC 1959
QY 2356 CATGACAGGTGCGCTTTCGACGGCTTGTGATGATGAATGAACAGAGCTTCCAACTTCATC 2415
Db 1960 CATGACAGGTGCGCTTTCGACGGCTTGTGATGATGAATGAACAGAGCTTCCAACTTCATC 2019
QY 2416 AAGAGCTCTGAGGACGGCTGCCCAACATAGCTGGAGAACCCACCTTACGTGCGCTGGG 2475
|||||

2020 AGAGGCTCTGAGGACGGGTGCGCCCAACAAATAGCTGGAGAACCCACCTTACGTGCGCTGGG 2079
QY 2476 GTGGTTGGGGGACCTCTCAGGCGGCCACCACTCTGTGCCTCCAGGACCAACAGTTTCTCTCC 2535
Db 2080 GTGGTTGGGGGACCTCTCAGGCGGCCACCACTCTGTGCCTCCAGGACCAACAGTTTCTCTCC 2139
QY 2536 ACACACTACAACTCTGAGAACTCTACGGCTTGAACGAGCCATCGCTCCACAGGGCG 2595
Db 2140 ACACACTACAACTCTGAGAACTCTACGGCTTGAACGAGCCATCGCTCCACAGGGCG 2199
QY 2596 CTGGTGAAGGCTCGGGGACACGCCCATTTGTGATCTCCCGCTCGACCTTTCTGCTGGCCAC 2655
Db 2200 CTGGTGAAGGCTCGGGGACACGCCCATTTGTGATCTCCCGCTCGACCTTTCTGCTGGCCAC 2259
QY 2656 GCGCATATACCGCGGACACTGAGACGGGGACGTGTGGAGCTCTCTGGGAGCAGCTCGCTCC 2715
Db 2260 GCGCATATACCGCGGACACTGAGACGGGGACGTGTGGAGCTCTCTGGGAGCAGCTCGCTCC 2319
QY 2716 TCCGTGCCAGAAATCCTGAGTTTAACTGTGTGGGGTGCTCTGTGTGGGGCCGACGTC 2775
Db 2320 TCCGTGCCAGAAATCCTGAGTTTAACTGTGTGGGGTGCTCTGTGTGGGGCCGACGTC 2379
QY 2776 TGGGGTTCCTGGGCAACACCTCAGAGGAGCTGTGTGGCTGGACCCAGCTGGGGCC 2835
Db 2380 TGGGGTTCCTGGGCAACACCTCAGAGGAGCTGTGTGGCTGGACCCAGCTGGGGCC 2439
QY 2836 TTCTACCCCTTCATGCGGAACCAACAGACCTGTCTCAGTCTGCCCCAGAGCCGTACACG 2895
Db 2440 TTCTACCCCTTCATGCGGNAACCAACAGCTGTCTCAGTCTGCCCCAGAGCCGTACACG 2499
QY 2896 TTCTACCGAGCCGGCCAGAGCCATGAGGAAGGCTCTACCTGTGGCTTACGACCTCTC 2955
Db 2500 TTCTACCGAGCCGGCCAGAGCCATGAGGAAGGCTCTCACCTGTGGCTTACGACCTCTC 2559
QY 2956 CCGCACCTCTACACTGTTCCACAGGCCACGCTCGGGGGGAGACCGTGGCCCGGGCC 3015
Db 2560 CCGCACCTCTACACTGTTCCACAGGCCACGCTCGGGGGGAGACCGTGGCCCGGGCC 2619
QY 3016 CTCTTCTGGAGTTCCTCCCAAGGACTCTAGCACTGTGGACCTCAGCTCTCTGTGG 3075
Db 2620 CTCTTCTGGAGTTCCTCCCAAGGACTCTAGCACTGTGGACCTCAGCTCTCTGTGG 2679
QY 3076 GGGGAGGCGCTGCTCATCAACCCAGTGTCTCAGGCGGGAAGCCGGAAGTGAAGTGA 3135
Db 2680 GGGGAGGCGCTGCTCATCAACCCAGTGTCTCAGGCGGGAAGCCGGAAGTGAAGTGA 2739
QY 3136 TTCCCTTGGGACATGTAGACCTGACAGCGGTCCAAATAGAGCCCTTTCGACGCTC 3195
Db 2740 TTCCCTTGGGACATGTAGACCTGACAGCGGTCCAAATAGAGCCCTTTCGACGCTC 2799
QY 3196 CCACCCCACTGACGCTCCCGTGAGCGAGCCATCCACAGCGAGGGGAGTGGGTGACG 3255
Db 2800 CCACCCCACTGACGCTCCCGTGAGCGAGCCATCCACAGCGAGGGGAGTGGGTGACG 2859
QY 3256 CTGCGGGCCCTCTGGAACCAATCAACGTCACCTCCGGGCTGGGTATCATATCCCTG 3315
Db 2860 CTGCGGGCCCTCTGGAACCAATCAACGTCACCTCCGGGCTGGGTATCATATCCCTG 2919
QY 3316 CAGGGCCCTGGCTCACAACCAACAGAGTCCCGGACGAGCCCATGCGCTGGCTGGCC 3375
Db 2920 CAGGGCCCTGGCTCACAACCAACAGAGTCCCGGACGAGCCCATGCGCTGGCTGGCC 2979
QY 3376 CTAAACAGGGTGGAGAGGCGCCGAGGGAGCTGTCTTGGGACGATGGAGAGCTTGGAA 3435
Db 2980 CTAAACAGGGTGGAGAGGCGCCGAGGGAGCTGTCTTGGGACGATGGAGAGCTTGGAA 3039
QY 3436 GTGCTGGAGCGAGGGCCCTTACACAGGTCTCTCTGGCCAGGAATTAACAGATCGTG 3495
Db 3040 GTGCTGGAGCGAGGGCCCTTACACAGGTCTCTCTGGCCAGGAATTAACAGATCGTG 3099
QY 3496 AATGAGCTGGTACGTGTGACCAAGTGGAGGCTGGCTGACGTGCAAGAGGTGACTGTC 3555
Db 3100 AATGAGCTGGTACGTGTGACCAAGTGGAGGCTGGCTGACGTGCAAGAGGTGACTGTC 3159

QY 3556 CTGGGCGTGGCCACAGCGCGCCAGAGGTCCTCTCAACGGTGTCCCTGTCTCCAACTTC 3615
DB 3160 CTGGGCGTGGCCACAGCGCGCCAGAGGTCCTCTCAACGGTGTCCCTGTCTCCAACTTC 3219
QY 3616 ACTACAGCCCGACACACCAAGTCTCTGGACATCTGTCTGTCTGTGATGGAGAGCAG 3675
DB 3220 ACTACAGCCCGACACACCAAGTCTCTGGACATCTGTCTGTCTGTGATGGAGAGCAG 3279
QY 3676 TTCTCGTCAGCTGGTGT 3694
DB 3280 TTCTCGTCAGCTGGTGT 3298

RESULT 2

US-09-949-016-3656
; Sequence 3656, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3656
; LENGTH: 3847
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-3656

Query Match 71.4%; Score 2644.6; DB 4; Length 3847;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2650; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 1036 GAAACCGGTGACACACCCCGCGCTCCAGAGAGTGCCACACAGTGCAGCGTCCCGCCC 1095
DB 639 GATGCCAGGACACACCCCGCGCTCCAGAGAGTGCCACACAGTGCAGCGTCCCGCCC 698
QY 1096 AACAGCGCTTGCATTTGCGCCCTGACAGGCGCATCACCCAGAACAGTGCAGGCGCGC 1155
DB 699 AACAGCGCTTGCATTTGCGCCCTGACAGGCGCATCACCCAGAACAGTGCAGGCGCGC 758
QY 1156 GGCTGCTGTACATCCCTGCAAGCAGGCGCTGCAGGAGCCAGATGGGCGAGCCCTGG 1215
DB 759 GGCTGCTGTACATCCCTGCAAGCAGGCGCTGCAGGAGCCAGATGGGCGAGCCCTGG 818
QY 1216 TGCCTTCTCCACCCAGCTACCCAGCTACAGCTGAGAACCTGAGCTCCTCTGAATG 1275
DB 819 TGCCTTCTCCACCCAGCTACCCAGCTACAGCTGAGAACCTGAGCTCCTCTGAATG 878
QY 1276 GGCTACAGGCGCACCTGTACCCGTACCCACCCCGACCTTTCCCGAAGACATCTGTACC 1335
DB 879 GGCTACAGGCGCACCTGTACCCGTACCCACCCCGACCTTTCCCGAAGACATCTGTACC 938
QY 1336 CTGGCGCTGAGATGATGATGAGACTGAGAACCCGCTCCACTTACGATCAAGATCCA 1395
DB 939 CTGGCGCTGAGATGATGATGAGACTGAGAACCCGCTCCACTTACGATCAAGATCCA 998
QY 1396 GCTAACAGGCGGTACAGAGTGCCTTTGGAGACCCCGCTGTCCACAGCGGCGACCGTCC 1455
DB 999 GCTAACAGGCGGTACAGAGTGCCTTTGGAGACCCCGCTGTCCACAGCGGCGACCGTCC 1058
QY 1456 CCACCTTACAGGCTGAGTCTCCGAGAGGCCCTTCGGGGTGATCGTGACCCGCGAGCTG 1515

DB 1059 CCACCTTACAGCGTGGAGTTCTCCGAGGAGCCCTTCGGGGTGATCGTGACCGGAGCTG 1118
QY 1516 GACGCGCGTGTCTGTGAACACGACGCGTGGCGCCCTGTCTTTTCGGAACGAGTTCCCT 1575
DB 1119 GACGCGCGTGTCTGTGAACACGACGCGTGGCGCCCTGTCTTTTCGGAACGAGTTCCCT 1178
QY 1576 GAGCTGTCCACCTCGCTGCGCTCGCAGTATATACAGGCGCTCGCGAGCACCTCAGTCCC 1635
DB 1179 GAGCTGTCCACCTCGCTGCGCTCGCAGTATATACAGGCGCTCGCGAGCACCTCAGTCCC 1238
QY 1636 CTGATGCTCAGCACCGAGCTGGACAGGATCACCTGTGGAAACGGGACCTTGGCGCCACG 1695
DB 1239 CTGATGCTCAGCACCGAGCTGGACAGGATCACCTGTGGAAACGGGACCTTGGCGCCACG 1298
QY 1696 CCCGCTGCGAACCTCTACCGGCTCTCACCCCTTTCTAACCCTGCGCTGAGAGACGCGGCTG 1755
DB 1299 CCCGCTGCGAACCTCTACCGGCTCTCACCCCTTTCTAACCCTGCGCTGAGAGACGCGGCTG 1358
QY 1756 GCACAGGCGTGTCTCTGCTAAACAGCAATGCATGGATGTGGTCTGCGAGCGGAGCCCT 1815
DB 1359 GCACAGGCGTGTCTCTGCTAAACAGCAATGCATGGATGTGGTCTGCGAGCGGAGCCCT 1418
QY 1816 GCCCTTAGCTGAGGTGCACAGGTGGATCCTGGATGTCTACATCTTCTGGGCCAGAG 1875
DB 1419 GCCCTTAGCTGAGGTGCACAGGTGGATCCTGGATGTCTACATCTTCTGGGCCAGAG 1478
QY 1876 CCCAAGAGCGTGGTGCAGCAGTACCTGGACGTTGTGGGATACCCGTTCTATCGCGCCATAC 1935
DB 1479 CCCAAGAGCGTGGTGCAGCAGTACCTGGACGTTGTGGGATACCCGTTCTATCGCGCCATAC 1538
QY 1936 TGGGCGCTGGGCTTCCACTGTGCGCTGGGCGTACTCTCCACCGCTATACCCGCGCAG 1995
DB 1539 TGGGCGCTGGGCTTCCACTGTGCGCTGGGCGTACTCTCCACCGCTATACCCGCGCAG 1598
QY 1996 GTGGTGGAGAACATGACAGGCGCCACTTCCCGCTGGAGCGTCCAAATGGAAACGACCTGGAC 2055
DB 1599 GTGGTGGAGAACATGACAGGCGCCACTTCCCGCTGGAGCGTCCAAATGGAAACGACCTGGAC 1658
QY 2056 TACATGGACTCCCGAGGAGCTTCAAGTTCAACAAAGATGGCTTCCCGGAGCTTCCCGGCGC 2115
DB 1659 TACATGGACTCCCGAGGAGCTTCAAGTTCAACAAAGATGGCTTCCCGGAGCTTCCCGGCGC 1718
QY 2116 ATGTGTCAGGAGCTGCAACAGGCGCGCGCTACATGATGATGATGATGATGATGATGATG 2175
DB 1719 ATGTGTCAGGAGCTGCAACAGGCGCGCGCTACATGATGATGATGATGATGATGATGATG 1778
QY 2176 AGCAGCTGGGCGCTTCCCGGAGCTACAGGCGCTACAGGCGCTTCCCGGAGGCGGCTT 2235
DB 1779 AGCAGCTGGGCGCTTCCCGGAGCTACAGGCGCTTCCCGGAGGCGCTTCCCGGAGGCGGCTT 1838
QY 2236 TTTCAACCAACGAGACCGGCGCGCTGATTTGGGAGGATTTGGCGCGGCTTCCACTGCGC 2295
DB 1839 TTTCAACCAACGAGACCGGCGCGCTGATTTGGGAGGATTTGGCGCGGCTTCCACTGCGC 1898
QY 2296 TTTCCCGAGCTTCCAAACCCCAACAGCCCTTGGCGCTTGGTGGGAGGACATGATGATGATG 2355
DB 1899 TTTCCCGAGCTTCCAAACCCCAACAGCCCTTGGCGCTTGGTGGGAGGACATGATGATGATG 1958
QY 2356 CATGACAGGCTGCTTCCAAACCCCAACAGCCCTTGGCGCTTGGTGGGAGGACATGATGATGATG 2415
DB 1959 CATGACAGGCTGCTTCCAAACCCCAACAGCCCTTGGCGCTTGGTGGGAGGACATGATGATGATG 2018
QY 2416 AGAGGCTCTGAGGAGCGGCTGCGGCGCGCTGAGTGGAGAACCCACCTTACGTGCGCTGGG 2475
DB 2019 AGAGGCTCTGAGGAGCGGCTGCGGCGCGCTGAGTGGAGAACCCACCTTACGTGCGCTGGG 2078
QY 2476 GTGGTGGGCGGAGCCCTTCCAGGCGCGCCACTTCTGTGCTTCCAGGCGCACAGTTTCTCTCC 2535
DB 2079 GTGGTGGGCGGAGCCCTTCCAGGCGCGCCACTTCTGTGCTTCCAGGCGCACAGTTTCTCTCC 2138
QY 2536 ACACACTAACCTGACAAACCTCTACGCGCTGACCGAGGCGCATGCGCTTCCACAGGCGG 2595
DB 2139 ACACACTAACCTGACAAACCTCTACGCGCTGACCGAGGCGCATGCGCTTCCACAGGCGG 2198

```
QY 2596 CTGGTGAAGGCTCGGGGACACGCCCATTTGTGATCTCCGCTCGACCTTTGCTGGCCAC 2655
Db 2199 CTGGTGAAGGCTCGGGGACACGCCCATTTGTGATCTCCGCTCGACCTTTGCTGGCCAC 2258
QY 2656 GGCATATACCGCGGCACTGAGCGGGGACGTGTGGAGTCTCTGGAGCAGCTCGCTCC 2715
Db 2259 GGCATATACCGCGGCACTGAGCGGGGACGTGTGGAGTCTCTGGAGCAGCTCGCTCC 2318
QY 2716 TCGTGCCAGAAATCTGCAGTTAACTCTGCGGGGTGCTCTGGTGGGGCCGACGTC 2775
Db 2319 TCGTGCCAGAAATCTGCAGTTAACTCTGCGGGGTGCTCTGGTGGGGCCGACGTC 2378
QY 2776 TCGGGCTTCTGGGCAACCTCAGAGGAGTGTGTGCGCTGGACCCAGCTGGGGCC 2835
Db 2379 TCGGGCTTCTGGGCAACCTCAGAGGAGTGTGTGCGCTGGACCCAGCTGGGGCC 2438
QY 2836 TTCTACCTCTCATGCGGAACCAACAGCTGTCTAGTCTGCCCGAGAGCGGTACAGC 2895
Db 2439 TTCTACCTCTCATGCGGAACCAACAGCTGTCTAGTCTGCCCGAGAGCGGTACAGC 2498
QY 2896 TTCTACGAGCGCGCCAGCAGGCGCATGAGGAGGCGCTCACCTGCGCTACGACTCTCTC 2955
Db 2499 TTCTACGAGCGCGCCAGCAGGCGCATGAGGAGGCGCTCACCTGCGCTACGACTCTCTC 2558
QY 2956 CCCCACCTCTACACACTGTTCACACGAGGCCACGTCGCGGGGAGACCGTGGCCCGGCC 3015
Db 2559 CCCCACCTCTACACACTGTTCACACGAGGCCACGTCGCGGGGAGACCGTGGCCCGGCC 2618
QY 3016 CTCCTCTGGAGTTCCTCAAGACTTAGACCTGTGACCTGTGGACCAACAGCTCTGTGG 3075
Db 2619 CTCCTCTGGAGTTCCTCAAGACTTAGACCTGTGACCTGTGGACCAACAGCTCTGTGG 2678
QY 3076 GGGGAGCCCTCTCATATACCCAGTGTCTCAGGCGGGAGGCGCGAGTGTACTGCTAC 3135
Db 2679 GGGGAGCCCTCTCATATACCCAGTGTCTCAGGCGGGAGGCGCGAGTGTACTGCTAC 2738
QY 3136 TTCCCTCTGGGACATGTGACGACTGTGACAGCGGTGCCAATAGAGCCCTTGGCAGCCTC 3195
Db 2739 TTCCCTCTGGGACATGTGACGACTGTGACAGCGGTGCCAATAGAGCCCTTGGCAGCCTC 2798
QY 3196 CCAACCCCACTGTGAGCTCCCGTGTAGCCAGGCATCCACAGGAGGGGAGTGGGTGAGC 3255
Db 2799 CCAACCCCACTGTGAGCTCCCGTGTAGCCAGGCATCCACAGGAGGGGAGTGGGTGAGC 2858
QY 3256 CTGGCGGCCCCCTGGACACCATCAAGCTCCACCTCGGSGTGGGTACATCATCCCCCTG 3315
Db 2859 CTGGCGGCCCCCTGGACACCATCAAGCTCCACCTCGGSGTGGGTACATCATCCCCCTG 2918
QY 3316 CAGGGCCCTGGCTCAACCAACAGAGTCCCGCCAGCAGCCCATGGCCCTGTGCTGTGGCC 3375
Db 2919 CAGGGCCCTGGCTCAACCAACAGAGTCCCGCCAGCAGCCCATGGCCCTGTGCTGTGGCC 2978
QY 3376 CTAAACAGGGTGGAGAGCGCCGAGGGAGCTGTCTGGGACATGGAGAGAGCTGGAA 3435
Db 2979 CTAAACAGGGTGGAGAGCGCCGAGGGAGCTGTCTGGGACATGGAGAGAGCTGGAA 3038
QY 3436 GTGCTGGAGCGAGGGCCCTACACACAGGTCTCTCTCGCCAGGAAATAACACGATCGTG 3495
Db 3039 GTGCTGGAGCGAGGGCCCTACACACAGGTCTCTCTCGCCAGGAAATAACACGATCGTG 3098
QY 3496 AATGAGCTGGTATGATGACAGTGAAGGAGCTGGCTTCAGCTCGAGAGGTGCTCTCCACTTC 3555
Db 3099 AATGAGCTGGTATGATGACAGTGAAGGAGCTGGCTTCAGCTCGAGAGGTGCTCTCCACTTC 3158
QY 3556 CTGGGCGTGGCCACCGCGGCCCGCAGAGGTCTCTCCAGCGGTGCTCTCTCTCCACTTC 3615
Db 3159 CTGGGCGTGGCCACCGCGGCCCGCAGAGGTCTCTCCAGCGGTGCTCTCTCTCCACTTC 3218
QY 3616 ACTACAGCCCGACACCAAGGTCTCTGGACATCTGTGCTCTGCTGTGTGATGGAGAGAG 3675
Db 3219 ACTACAGCCCGACACCAAGGTCTCTGGACATCTGTGCTCTGCTGTGTGATGGAGAGAG 3278
```

```
QY 3676 TTTCTCGTCAGCTGGTGT 3694
Db 3279 TTTCTCGTCAGCTGGTGT 3297

RESULT 3
US-09-397-787-175
; Sequence 175, Application US/09397787
; Patent No. 6468758
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR OVARIAN
; TITLE OF INVENTION: CANCER THERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.466C2
; CURRENT APPLICATION NUMBER: US/09/397,787
; NUMBER OF SEQ ID NOS: 334
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 175
; LENGTH: 449
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-397-787-175

Query Match 10.0%; Score 370.6; DB 3; Length 449;
Best Local Similarity 98.9%; Pred No. 3.4e-67;
Matches 373; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 63 CGGATCTTACTCGCGGGAGAGAACCCAGCCCAAGCCGTCCCGAAACCGAGTCCGGAGA 122
Db 73 CGGCAAGTACTCGCGGGAGAGAACCCAGCCCAAGCCGTCCCGAAACCGAGTCCGGAGA 132
QY 123 GGAGTTCCGATCGAGAGTTGAACCCAGCTGTGGAGAGGCCCGGAGTGCATCTTCC 182
Db 133 GGAGTTCCGATCGAGAGTTGAACCCAGCTGTGGAGAGGCCCGGAGTGCATCTTCC 192
QY 183 TCCCGTAGGCTGGCGGAGCTCCAGCTGATCTGAAGATACAGGAGAGGAGCGAACTCGC 242
Db 193 TCCCGTAGGCTGGCGGAGCTCCAGCTGATCTGAAGATACAGGAGAGGAGCGAACTCGC 252
QY 243 CTGGAAGAACTAAAGCTTGACGGTTGAGCAAGATGGGAGAGAGGAGCGAGACTCAT 302
Db 253 CTGGAAGAACTAAAGCTTGACGGTTGAGCAAGATGGGAGAGAGGAGCGAGACTCAT 312
QY 303 AGCGAACCTCAATGTCTATCTTGGCCCAAGTATGCTTGGAGCGGAGAGGAGCGCTCGCA 362
Db 313 AGCGAACCTCAATGTCTATCTTGGCCCAAGTATGCTTGGAGCGGAGAGGAGCGCTCGCA 372
QY 363 GGTGACCACTCACTCCCTCAGTGGCACCCAGGAGACCGGCTGGATGACCCCGAGCTGGA 422
Db 373 GGTGACCACTCACTCCCTCAGTGGCACCCAGGAGACCGGCTGGATGACCCCGAGCTGGA 432
QY 423 AAAGCTGTGGCACAAAG 439
Db 433 AAAGCTGTGGCACAAAG 449
```

```
RESULT 4
US-08-430-925A-3
; Sequence 3, Application US/08430925A
; Patent No. 5763252
; GENERAL INFORMATION:
; APPLICANT: Skadsen, Ronald W
; APPLICANT: Tibbot, Brian K
; TITLE OF INVENTION: Cloned Alpha-Glucosidase from Barley
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
```


/ STATE: WI
/ COUNTRY: US
/ ZIP: 53703
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/430,925A
/ FILING DATE:
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Seay, Nicholas J
/ REGISTRATION NUMBER: 27,386
/ REFERENCE/DOCKET NUMBER: 960296.93171
/ TELEPHONE: 608-251-5000
/ TELEFAX: 608-251-9166
/ INFORMATION FOR SEQ ID NO: 3:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2752 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 42..2675
/ US-08-430-925A-3

Query Match 10.0%; Score 369.8; DB 1; Length 2752;
Best Local Similarity 53.6%; Pred. No. 8.1e-67;
Matches 966; Conservative 0; Mismatches 772; Indels 63; Gaps 7;

QY	1687	GGCCACAGCCCGGTGGAAACCTCTACGGGTCTCACCCCTTCTACCTGGGGTGGAGGAC	1746
DB	648	GGCTCTACGTGGACGTCAACCTCTACGGCTGACCCCTTCTACATGACGTGGGGCA	707
QY	1747	GGGGGTGGGACACAGGGGTGTTCTGCTAAACAGCAATGCCATGATGTGGTCTGACG	1806
DB	708	CGGGGACCGGACGCGGTGCTCTCTGCTCAGTAGGACAGGATGACGCTG---CTCTAC	764
QY	1807	CCGAGCCCTGCCCTTAGCTGAGGTGCGACAGGTGGGATCTGGAGTCTACATCTTCCTG	1866
DB	765	GGCGGTCTCTACGTCACTACAGGTCTATCGGGGGCTCTCGACTTCTACTTCTTCGCC	824
QY	1867	GGCCAGAGCCCAAGAGCGTGTGACGAGTACTGGAAGTGTGGGATACCGGTTCTATG	1926
DB	825	GGCCCAACCCCTCGCCGTGCTGCGACAGTACACCCAGTCTATCGCCCGCTGCCCA	884
QY	1927	CGCCATACCTGGGGCTGGGCTTCCACCTGTGCGGTGGGGTACTCTCCACCGCTATC	1986
DB	885	ATGCGGTACTGTGCTTTCGGGTTCACCAAGTCCGGTACCGGTACTGAACTGTCTGAC	944
QY	1987	ACCCGCAAGTGTGGAGAACATGACAGGGGCCCACTTCCCTCGGACGTCCAATGGAAC	2046
DB	945	CTGGAGCGTGTGTGGCCCGGATACGCCAAGGCCCGGATCCCGCTGGAGTGTGACCC	1004
QY	2047	GACCTGGACTATGAGACTCCCGGAGGACTTCACTTCAACAGGATGGCTT-----C	2100
DB	1005	GATATCGACTATGACCGGGTTCAGGACTTCACCTTGGACCGGTGTCACTTCAACCGC	1064
QY	2101	CGGGACTTCCCGCCATGTGAGGAGTGTGACAGGGCGCGGCTACATGATGATC	2160
DB	1065	GCGAGCTCCGGCGGTTCGTGACCGGCTTACCGGAACGCCAGAAATACGTCCTCATC	1124
QY	2161	GTGGATCTCGCACTCAGCAGCTCGGGCCCTCGCGGAGCTACAGGCCCTACGACGAGGT	2220
DB	1125	CTAGACCCAGGATCCGGGTGGACCCCATCGACGCGACGTACGGGACGTTGTCGCCGG	1184
QY	2221	CTGCGAGGGGGTTCATTCACCAACGAGACCGGCCAGCGGCTGATGTGGGAAGGTATGG	2280

DB	1185	ATGACGAGGACATCTTCTCT---GAAGCGGAACGGCAAAACTTCTGTCGGCAACGTGTGG	1241
QY	2281	CCCGGGTCCACTGCTTCCCGACTTACCAACCCCAAGCCCTGGCTGGTGGGAGGAC	2340
DB	1242	CCGGGCGACGTCTACTTCCCGGACTTCATGCAACCCAGCCGCGCGAGTCTTGGGCGCG	1301
QY	2341	ATGGTGGGTGAGTTTCATGACAGGTGCTTTCGAGGGCTTGTGGATTGACATGAACGAG	2400
DB	1302	GAGATCTCCCTCTTCCGCGGACCAATCCCGGTGCGAGGGCTGTGGATCGACATGAACGAG	1361
QY	2401	CTTCCAACTTCATCAGAGGCTCTGAGGACGGTCTGCCCAACAAATGAGCTGAGAGAACCA	2460
DB	1362	ATCTCCAACTTCTACA---ACCGGAGCCATGAAGCGCTCGACACCCGCTACCGG	1418
QY	2461	CCCTAGTGCCTGGGGTGTGGGGGACCTTCAAGGCGGCCACCAATCTGTGTCTCCAGC	2520
DB	1419	ATCAACAACGACGGGACGGGCGGCCCATCAACAACAAGACGGTGGCGCGCTGCGCGTG	1478
QY	2521	GACCAAGTTTCTCTCACACACTACAACCTGCAACCTTACGGCTGACCGNAGCCATC	2580
DB	1479	CATACGGGGGGTCAACCGAGTACGAGGAGCAACCTCTTCCGGCTCTCTCGAGGGCGCG	1538
QY	2581	GCCTCCACACAGGGCGCTGGTGAAGGCTCGGGGACACCGCCCATTTGTGATCTCCCGCTG	2640
DB	1539	GCACCGGGCGGGGTGCTGAGGACACCGCGCGGCCCTTCTGCTCAGCAGGTCC	1598
QY	2641	ACCTTTGTGCGCCACGGCCGATACGCGGCCCATGGAACGGGGGACGTGTGGAGCTCTCG	2700
DB	1599	ACCTTGTGGGCTCGGGCGCTACACCGCTTACTGGACTGGGACAAACGCGCAACGTGG	1658
QY	2701	GAGCAGCTCGCTCTCGTGCAGAAATCTGTCAGTTTAACTGTGTGGGGTGGCTCTG	2760
DB	1659	GGCGACCTCGCTACTCTCAATCAACACCATGCTCAGCTTCGGCTCTCTCGGCAATGCCATG	1718
QY	2761	GTGCGGGCGGAGCTGTGGGGCTTCTGGGCAACACCTCAGAGAGCTGTGTGTCGCTGG	2820
DB	1719	ATCGGGCGCGCATCTGCGGGTTCAAGCGCAACAACAGAGAGAGCTTCCGGTTCGGTGG	1778
QY	2821	ACCCAGCTGGGGGCTTCTACCCCTTTCATGCGGAACCAACAGCCCTGCTAGTCTGCC	2880
DB	1779	ATCCAGCTCGGAGCTTCTACCCCTTCTCGAGGACCACTCGGCGATCTTACCGTCCGG	1838
QY	2881	CAGGACCGGTACAGCTTACGAGCGCGGCCAGAGGCCATAGAGNAGGCCCTCACCTG	2940
DB	1839	CGAGAGTTGTA---CCTGTGGCGCTGCTGGGGGCTCGGGCAGGAAGCGCTCGGGCTC	1895
QY	2941	CGCTACGCACTCTCCCGCACCTTACACACTGTTCCACAGGCCCAACGTCGGGGGGAG	3000
DB	1896	CGGTACAGCTCTCCCTTACTTCTACAGCTCATGTACAGGCGCAATGACGGGGGCG	1955
QY	3001	ACCGTGGCGCGCCCTCTTCTCTGGAGTTCCCAAGGACTCTAGCACCTGGACTGTGGAC	3060
DB	1956	CCAAATCGCGCGCGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	2015
QY	3061	CACAGCTCTCTGTGGGGAGGCGCTGTCTATCAACCCAGTGTCTCAGGCGCGGGAAGGCC	3120
DB	2016	AGACAGTTCTCTCTCGCGCGGGTCTCTGTTTCTGCGGTGCTCAGGCGCGGCGCGACA	2075
QY	3121	GAGTGTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	3180
DB	2076	ACCGTCGACGCTACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	2128
QY	3181	GCCCTTGGCAGCTCTCCACCCCACTGTCAGCTCTCCCGTGAGCGACCATCCACAGCGAG	3240
DB	2129	CTTCCGCTGCGCAACCGGACCGGCAACGACGCTC-----	2162
QY	3241	GGGCAAGTGGTGAACGCTGCGGCCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	3300
DB	2163	-----AGGCTGCG	2210
QY	3301	TACATCATCTCCCTCGAGGGCCCTGGCTTCAACCAACAGAGTCCCGCGCAGGACCCATG	3360
DB	2211	ACCATCTCTCCCTGACGAGCGCGTGTACTGTCGCGCGCGCGCGCGCGCGCGCGCGCTC	2270

Qy	3361	GCCTGGCTGTGGCCCTAACCAAGGGTGGAGAGGCCCGAGGGAGCTGTCTCTGGACGAT	3420
Db	2271	CACCTCCTGGTCCGCTCGCGAGACGGGACGGCCAGCGGCTACCTTTCTCTGGACGAC	2330
Qy	3421	GGAGAGCCTGGAAGTGTGGAGCGAGGGGCGCTACACAGGTCATCTTCTCTGGCCAGG	3480
Db	2331	GGCGACTCTCCGGATGTGGCAGGAGAAGCGATTGGAGCATGGTAAGTTCAACTACAAG	2390
Qy	3481	A	3481
		—	
Db	2391	A	2391

RESULT 5

```

US-09-949-016-11765
; Sequence 11765, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11765
; LENGTH: 22404
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(22404)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-11765

```

Query Match	9.2%;	Score 340.4;	DB 4;	Length 22404;	
Best Local Similarity	98.3%;	Pred. No. 1.7e-60;			
Matches 344;	Conservative 0;	Mismatches 6;	Indels 0;	Gaps 0;	
QY	1036	GA AACGGTGCACACCCCGCGCTCCAGAGCAGTGC	CCACACAGTGC	GACGTCCCGCCCC	1095
DB	5303	GATGCCCAGGCAACCCCGCGCTCCAGAGCAGTGC	CCACACAGTGC	GACGTCCCGCCCC	5362
QY	1096	AACAGCCGCTTCGATTGCGGCCCTTGACAAGGCC	CCATCACCCAGGAACAGTGC	GAGGCCCGC	1155
DB	5363	AACAGCCGCTTCGATTGCGGCCCTTGACAAGGCC	CCATCACCCAGGAACAGTGC	GAGGCCCGC	5422
QY	1156	GGCTGTCTACATCTCTGCAAGACAGGGGCTG	CAGGGAGCCAGATG	GGGGCAGCCCTTG	1215
DB	5423	GGCTGTCTACATCTCTGCAAGACAGGGGCTG	CAGGGAGCCAGATG	GGGGCAGCCCTTG	5482
QY	1216	TGCTTCTTCCACCCAGCTACCCGAGTCAAGCTG	GAGAACCTGAGTCTCCTCTG	GAATG	1275
DB	5483	TGCTTCTTCCACCCAGCTACCCGAGTCAAGCTG	GAGAACCTGAGTCTCCTCTG	GAATG	5542
QY	1276	GGCTACACGGCCACCTTGACCCCGTACCACCC	CCACCTCTTTCCCCCAAGGACATCTCTG	AGCC	1335
DB	5543	GGCTACACGGCCACCTTGACCCCGTACCACCC	CCACCTCTTTCCCCCAAGGACATCTCTG	AGCC	5602
QY	1336	CTCGGGCTGACGTGATGTAGTGAGACTGAGAA	CCGCGCTCCACTTCACGAT		1385
DB	5603	CTCGGGCTGACGTGATGTAGTGAGACTGAGAA	CCGCGCTCCACTTCACGAT		5652

RESULT 6

US-09-949-016-15398
; Sequence 15398, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15398
; LENGTH: 22404
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(22404)
; OTHER INFORMATION: n = A,T,C or G

US-09-949-016-15398

Query Match 9.2%; Score 340.4; DB 4; Length 22404;
Best Local Similarity 98.3%; Pred.No. 1.7e-60;
Matches 344; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

	QY	DB	QY	DB	QY	DB	QY	DB	QY	DB	QY	DB	QY	DB	QY	DB	
1036	GAAACGGGTGCACACCCCGGCTCCAGAGCAGTGCACACACAGTGCGAGCTGCCCCCC	1095	5303	GATGCCACGACACACCCCGGCTCCAGAGCAGTGCACACACAGTGCGAGCTGCCCCCC	5362	1096	AACAGCGCTTCGATTGCGCCCCTGACAAGGCCATCAACCAGGAACAGTGCAGGCGCCGC	1155	5363	AACAGCGCTTCGATTGCGCCCCTGACAAGGCCATCAACCAGGAACAGTGCAGGCGCCGC	5422	1156	GGCTGCTGTACTATCCCTGCAAGACGGGGCTGCAGGGGCCAGATGGGGCAGCCCTGG	1215	5423	GGCTGCTGTACTATCCCTGCAAGACGGGGCTGCAGGGGCCAGATGGGGCAGCCCTGG	5482
1216	TGCTTTCTCCACCCAGCTACCCAGCTCAAGCTGGAGAACCTGAGCTCTCTGAAATG	1275	5483	TGCTTTCTCCACCCAGCTACCCAGCTCAAGCTGGAGAACCTGAGCTCTCTGAAATG	5542	1276	GGCTACAGGCGCACCTGACCGGTACACACCCACCTCTTCCCAAGGACATCTGACC	1335	5543	GGCTACAGGCGCACCTGACCGGTACACACCCACCTCTTCCCAAGGACATCTGACC	5602	1336	CTGGGGCTGGACGTGATGATGGAGACTGAGAAGCGCTCCACTTCACGAT	1385	5603	CTGGGGCTGGACGTGATGATGGAGACTGAGAAGCGCTCCACTTCACGAT	5652

RESULT 7
US-09-744-926-1
; Sequence 1, Application US/09744926
; Patent No. 6794558
; GENERAL INFORMATION:
; APPLICANT: FROBERG, Claus
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING AN ALPHA-GLUCOSIDASE, PLANT
; TITLE OF INVENTION: SYNTHESIZE A MODIFIED STARCH, THE GENERATION OF THE PLANT
; TITLE OF INVENTION: MODIFIED STARCH
; FILE REFERENCE: 514413-3865
; CURRENT APPLICATION NUMBER: US/09/744,926
; CURRENT FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/EP99/05536
; PRIOR FILING DATE: 1999-07-31
; PRIOR APPLICATION NUMBER: DE 198 36 097.5
; PRIOR FILING DATE: 1998-07-31


```
QY 2333 GGGAGGACATGGTGGCTGAGTTCCATGACACAGGTGCGCTTCGACGCTTGTGGATTGACA 2392
Db 61 GGGCAAAATGAATGACGATATTTTCCATCAAGAAGTGCAATATGATGACTTTGGATTGACA 120
QY 2393 TGAACGAGCTTCCAACTTCATCAGAGGCTCTGAGGACGGCTGCCCAACAATGAGCTGG 2452
Db 121 TGAATGAGTTTCCAGCTTATTCAGGTTCAACAAAGGATGTAATGTAACAATGGA 180
QY 2453 AGAACCCACCTACGTCGCTGGGGTGGTGGGGGACCCCTCCAGGGGGCCACCATCTGTG 2512
Db 181 ATTATCCACGGTTTACTCTGTATATCTTTCGACAACTCATGTATTCCTCAAAACAATTTGCA 240
QY 2513 CTTCCAGCACACAGTTTCTTCACACACTCAACCTGACAACTCTACGGCTGACGG 2572
Db 241 TGGATCTGTGCAGAACTGGGGTAAACAGTATGTTTCATAGCTCTATGATACAGCA 300
QY 2573 AAGCCATCCCTCCACAGAGGGCTGGTGAAGS---CTCGGGGGACACCCCATTTGTA 2629
Db 301 TGGCTATAGCCACAGAGCAAGCTGTACAAAAAGTTTTCCTAATAGAAAGCTTCATTC 360
QY 2630 TCTCCGCTCGACCTTTGTGTCGCCACCGCGGATACCGCGGCCACTGGACGGGGGACGTGT 2689
Db 361 TTACCGCTCAACATTTGTGATCTGGAAGACATGCTCTCATTTGGTTAGGACAAATA 420
QY 2690 GGAGTCTCTGGAGACAGCTGCGCTCTCCGTGCGCAGAAATCTGCAAGTTTAACTGCTGG 2749
Db 421 CTGCTTTCATGGGAAACAAATGGAATGCTCTATAACTGGAATGCTGGAAGTTTCAGTTGTTG 480
QY 2750 GGGTGCTCTGTGGGGCGGACGCTGCGGCTTCTCTGGGCAACACCTCAGAGGAGCTGT 2809
Db 481 GAATACCTTTGGTGAGCAGACATCTGTGGAATTTGTGCTGAAACCCACAGAAAGACTTT 540
QY 2810 GTGTGCGCTGGACCCAGCTGGGGGCTCTACCCCTTCATGCGGAACCAACAGCCTGC 2869
Db 541 GCAGAGATGGATGCAACTTGGGGCATTTTATCCATTTTCCAGAAACCATATTTCTGAGC 600
QY 2870 TCAGTCTGCCAGGAGCGGTACAGCTTCAGGAGCGGGCCAGCAGG-----CCATGA 2923
Db 601 GATATGAACATCAGGATCTGCAATTTTGGGCGAATTCACCTTTGGTTAAATCATCAA 660
QY 2924 GGAAGCCCTCACCTCGCGCTAGCACTCTCCGCCACTCTACACACTGTTCACCCAGG 2983
Db 661 GCAGATTTAACTATTCGCTACACCTTATACCTTCTCTACACTCTGTTTATAAG 720
QY 2984 CCCAGCTCGGGGGAGACCGTGGCGCGGCTCTTCTCTGGAGTTCCCAAGGACTCTA 3043
Db 721 CCATGTGTTGGAGAAAAGTAGCAAGACCACTTCTCATGAGTTTATGAGGATACGA 780
QY 3044 GCACCTGAGCTGTGGACCAAGCTCTCTGTGGGGAGGCGCTCTCATCACCAGTGC 3103
Db 781 ACAGCTGGATTGAGGACACTGAGTTTGTGGGGCGCTCATTTACTTACTCTCTGTTTC 840
QY 3104 TCCAGCGCGGAGCGGAGTGACTGCTGCTCTTCCCTTGGGCAATGTTAGCA 3158
Db 841 TAAACAGGGAGCAGATATCTGTAGTGCTTACATCCCTCATGCTATTTGGTATGA 895

RESULT 9
US-09-949-016-18094
; Sequence 18094, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18094
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-18094
Query Match 4.7%; Score 175.2; DB 4; Length 601;
Best Local Similarity 77.8%; Pred. No. 7.6e-27;
Matches 210; Conservative 1; Mismatches 59; Indels 0; Gaps 0;

QY 3306 CATCCCTTCGAGGCGCTTGGCTTCAACACACAGAGTCCCGCCAGCAGCCCATGGCCCT 3365
Db 217 CTTCCCTTTCCAGGCGCTTGGCTTCAACACACAGAGTCCCGCCAGCAGCCCATGGCCCT 276
QY 3366 GCGTGTGGCCCTTAAACAAAGGTGGAGAGCCCGAGGGAGCTGTTCTGGGACGATGGAGA 3425
Db 277 GCGTGTGGCCCTTGAACAAAGGTGGAGAGCCCGAGGGAGCTGTTCTGGGACGATGGAGA 336
QY 3426 GAGCCTGGAAGTGTGGAGCGAGGGGCTTACACACAGAGTCTTCTGCGCCAGGAATAA 3485
Db 337 GAGCCTGGAAGTGTGGAGCGAGGGGCTTACACACAGAGTCTTCTGCGCCAGGAATGT 396
QY 3486 CACGATCTGTAATGAGCTGTACGTGTACCACTGAGGGAGCTGCGCTTCAGCTGCAGAA 3545
Db 397 GAGTCTGGGGCTGCTCAGGCTGTGTGGGCGAGGGCGCGCTCGGGGTTCAGAAAGGGGTGAG 456
QY 3546 GTGAGTCTCTGGGCGTGGCCACCGCGCC 3575
Db 457 GGGACCTGGGCTTGGGGGTCCACCATGGC 486

RESULT 10
US-09-949-016-128902
; Sequence 128902, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 128902
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-128902
Query Match 4.7%; Score 175.2; DB 4; Length 601;
Best Local Similarity 77.8%; Pred. No. 7.6e-27;
Matches 210; Conservative 1; Mismatches 59; Indels 0; Gaps 0;

QY 3306 CATCCCTTCGAGGCGCTTGGCTTCAACACACAGAGTCCCGCCAGCAGCCCATGGCCCT 3365
Db 217 CTTCCCTTTCCAGGCGCTTGGCTTCAACACACAGAGTCCCGCCAGCAGCCCATGGCCCT 276
QY 3366 GCGTGTGGCCCTTAAACAAAGGTGGAGAGCCCGAGGGAGCTGTTCTGGGACGATGGAGA 3425
Db 277 GCGTGTGGCCCTTGAACAAAGGTGGAGAGCCCGAGGGAGCTGTTCTGGGACGATGGAGA 336
QY 3426 GAGCCTGGAAGTGTGGAGCGAGGGGCTTACACACAGAGTCTTCTGCGCCAGGAATAA 3485
```

```
QY 2333 GGGAGGACATGGTGGCTGAGTTCCATGACACAGGTGCGCTTCGACGCTTGTGGATTGACA 2392
Db 61 GGGCAAAATGAATGACGATATTTTCCATCAAGAAGTGCAATATGATGACTTTGGATTGACA 120
QY 2393 TGAACGAGCTTCCAACTTCATCAGAGGCTCTGAGGACGGCTGCCCAACAATGAGCTGG 2452
Db 121 TGAATGAGTTTCCAGCTTATTCAGGTTCAACAAAGGATGTAATGTAACAATGGA 180
QY 2453 AGAACCCACCTACGTCGCTGGGGTGGTGGGGGACCCCTCCAGGGGGCCACCATCTGTG 2512
Db 181 ATTATCCACGGTTTACTCTGTATATCTTTCGACAACTCATGTATTCCTCAAAACAATTTGCA 240
QY 2513 CTTCCAGCACACAGTTTCTTCACACACTCAACCTGACAACTCTACGGCTGACGG 2572
Db 241 TGGATCTGTGCAGAACTGGGGTAAACAGTATGTTTCATAGCTCTATGATACAGCA 300
QY 2573 AAGCCATCCCTCCACAGAGGGCTGGTGAAGS---CTCGGGGGACACCCCATTTGTA 2629
Db 301 TGGCTATAGCCACAGAGCAAGCTGTACAAAAAGTTTTCCTAATAGAAAGCTTCATTC 360
QY 2630 TCTCCGCTCGACCTTTGTGTCGCCACCGCGGATACCGCGGCCACTGGACGGGGGACGTGT 2689
Db 361 TTACCGCTCAACATTTGTGATCTGGAAGACATGCTCTCATTTGGTTAGGACAAATA 420
QY 2690 GGAGTCTCTGGAGACAGCTGCGCTCTCCGTGCGCAGAAATCTGCAAGTTTAACTGCTGG 2749
Db 421 CTGCTTTCATGGGAAACAAATGGAATGCTCTATAACTGGAATGCTGGAAGTTTCAGTTGTTG 480
QY 2750 GGGTGCTCTGTGGGGCGGACGCTGCGGCTTCTCTGGGCAACACCTCAGAGGAGCTGT 2809
Db 481 GAATACCTTTGGTGAGCAGACATCTGTGGAATTTGTGCTGAAACCCACAGAAAGACTTT 540
QY 2810 GTGTGCGCTGGACCCAGCTGGGGGCTCTACCCCTTCATGCGGAACCAACAGCCTGC 2869
Db 541 GCAGAGATGGATGCAACTTGGGGCATTTTATCCATTTTCCAGAAACCATATTTCTGAGC 600
QY 2870 TCAGTCTGCCAGGAGCGGTACAGCTTCAGGAGCGGGCCAGCAGG-----CCATGA 2923
Db 601 GATATGAACATCAGGATCTGCAATTTTGGGCGAATTCACCTTTGGTTAAATCATCAA 660
QY 2924 GGAAGCCCTCACCTCGCGCTAGCACTCTCCGCCACTCTACACACTGTTCACCCAGG 2983
Db 661 GCAGATTTAACTATTCGCTACACCTTATACCTTCTCTACACTCTGTTTATAAG 720
QY 2984 CCCAGCTCGGGGGAGACCGTGGCGCGGCTCTTCTCTGGAGTTCCCAAGGACTCTA 3043
Db 721 CCATGTGTTGGAGAAAAGTAGCAAGACCACTTCTCATGAGTTTATGAGGATACGA 780
QY 3044 GCACCTGAGCTGTGGACCAAGCTCTCTGTGGGGAGGCGCTCTCATCACCAGTGC 3103
Db 781 ACAGCTGGATTGAGGACACTGAGTTTGTGGGGCGCTCATTTACTTACTCTCTGTTTC 840
QY 3104 TCCAGCGCGGAGCGGAGTGACTGCTGCTCTTCCCTTGGGCAATGTTAGCA 3158
Db 841 TAAACAGGGAGCAGATATCTGTAGTGCTTACATCCCTCATGCTATTTGGTATGA 895

RESULT 9
US-09-949-016-18094
; Sequence 18094, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
```


Db 380 GTGACAGACTGCACACAGGCGCGCTACATGATGATCGTGTGTGTCGCCCCACA 439
QY 2179 AGCTGGGCCCTG 2191
Db 440 CTGTGGGTCTTTG 452

RESULT 14
US-09-949-016-128889
; Sequence 128889, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 128889
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-128889

Query Match 4.3%; Score 157.4; DB 4; Length 601;
Best Local Similarity 71.3%; Pred. No. 3.6e-23;
Matches 266; Conservative 1; Mismatches 17; Indels 89; Gaps 1;

QY 1908 TGTGGGATACCCGTTTCATCCCGCATACCTGCGGCTTGGGCTTCCACCTGTGCGGTGGG 1967
Db 80 TGCAGGATACCCGTTTCATCCCGCATACCTGCGGCTTGGGCTTCCACCTGTGCGGTGGG 139

QY 1968 CTACTCTCAGCGCTATCACCGCCAGTGTGTGGAGACATGACAGGCGCCACTTCCC 2027
Db 140 CTACTCTCAGCGCTATCACCGCCAGTGTGTGGAGACATGACAGGCGCCACTTCCC 199

QY 2028 CCT----- 2030
Db 200 CTGTGTAGTGGGTGGTGGCAGGGAGCGGCTGGCGGACCGCTCTCTCAG 259

QY 2031 -----GGAGCTCAATGGAAACGACCTGGACTAC 2058
Db 260 GCGCCAGCAGACGGTCCCGTGTGTGGCTGCAGGACGTCARTGGAAACGACCTGGACTAC 319

QY 2059 ATGGACTCCGGAGGACTTCAGTTCAACAGGATGGCTTCGGGACTTCGGGCGCATG 2118
Db 320 ATGGACTCCGGAGGACTTCAGTTCAACAGGATGGCTTCGGGACTTCGGGCGCATG 379

QY 2119 GTGCAGGAGCTGCACACAGGCGCGCGCTACATGATGATCGTGGATCTGCCATCAGC 2178
Db 380 GTGCAGGAGCTGCACACAGGCGCGCGCTACATGATGATCGTGGATCTGCCATCAGC 439

QY 2179 AGCTGGGCCCTG 2191
Db 440 CTGTGGGTCTTTG 452

RESULT 15
US-09-949-016-18077
; Sequence 18077, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18077
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-18077

Query Match 4.2%; Score 155; DB 4; Length 601;
Best Local Similarity 95.8%; Pred. No. 1.1e-22;
Matches 158; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1366 AACCGCTCCACTTCACGATCAAGATCCAGTAAACAGGCGCTACGAGTGCCCTTGGAG 1425
Db 287 AACCCCATCTCTTSTAGATCAAGATCCAGTAAACAGGCGCTACGAGTGCCCTTGGAG 346

QY 1426 ACCCGCGTGTCCACAGCGGCGCACCGTCCCACTCTACAGCGTGGAGTTCTCCGAGGAG 1485
Db 347 ACCCGCGTGTCCACAGCGGCGCACCGTCCCACTCTACAGCGTGGAGTTCTCCGAGGAG 406

QY 1486 CCCTTGGGGTGATCGTGCACCGGAGCTGGAGCGCGGCTGCTG 1530
Db 407 CCCTTGGGGTGATCGTGCACCGGAGCTGGAGCGCGGCTGCTG 451

Search completed: April 4, 2005, 16:02:28
Job time : 613 secs

ALIGNMENTS

121 GAGGAGTTCGCGATGAGAGGTTGAACAGCTGTGGGAGAAAGCCAGCGACTGCATCTT 180
181 CTTCCCGTGAGGCTGCGCGAGCTCCAGCTGATCTGAAGATACAGAGAGGAGCGAATC 240
181 CTTCCCGTGAGGCTGCGCGAGCTCCAGCTGATCTGAAGATACAGAGAGGAGCGAATC 240
241 GCCTGGAAGAAATATAAGCTTTGACGGCTTGGACGAAAGTGGGAGAAAGCGAGACTC 300
241 GCCTGGAAGAAATATAAGCTTTGACGGCTTGGACGAAAGTGGGAGAAAGCGAGACTC 300
301 ATACGCAACTCAATGTCTATCTTTGGCAAAGTATGTTCTGGA CGGAAGAAAGCGCTCGG 360
301 ATACGCAACTCAATGTCTATCTTTGGCAAAGTATGTTCTGGA CGGAAGAAAGCGCTCGG 360
361 CAGGTGACAGCAACTCCCTCAGTGCACCCAGGAGAGGGCTGATGACCCAGGCTG 420
361 CAGGTGACAGCAACTCCCTCAGTGCACCCAGGAGAGGGCTGATGACCCAGGCTG 420
421 GAAAAGCTGTGGCAAGGCGAAGACTCTGGGAAATTTCTCGGCGAAGAACTGGACAAG 480
421 GAAAAGCTGTGGCAAGGCGAAGACTCTGGGAAATTTCTCGGCGAAGAACTGGACAAG 480
481 CTTGCGGGAGTTCTGCAATCAAAAGAAAGTTTACAGATPACAAAGCTCTGTCTGGAG 540
481 CTTGCGGGAGTTCTGCAATCAAAAGAAAGTTTACAGATPACAAAGCTCTGTCTGGAG 540
541 ACCCTGAGCAGGACCGAAGAAATCCACGAGAGCTTATAGCCCTCGGACCTGAGCGAC 600
541 ACCCTGAGCAGGACCGAAGAAATCCACGAGAGCTTATAGCCCTCGGACCTGAGCGAC 600
601 ATCAAGGGCAGCGTCTGCAAGCAGGAGCACAAGGAGTGAAGAGAGCTGGCGAGCATC 660
601 ATCAAGGGCAGCGTCTGCAAGCAGGAGCACAAGGAGTGAAGAGAGCTGGCGAGCATC 660
661 AACGAGGCTGAGACCGCTGCGCAGGCTCAGCCACGAGGCTACAGCATGAGGCTGAG 720
661 AACGAGGCTGAGACCGCTGCGCAGGCTCAGCCACGAGGCTACAGCATGAGGCTGAG 720
721 TTGAGGAGCCAGGCTGATGACCTGTGGAGCTTGGCGAGTCCGCGAGTCCGCGAGCTC 780
721 TTGAGGAGCCAGGCTGATGACCTGTGGAGCTTGGCGAGTCCGCGAGTCCGCGAGCTC 780
781 AAGGAGCTGAGGCGTTCGCGAGGAGCTCAAGCACTTGAAGGCAAAATCGAAGAGCAC 840
781 AAGGAGCTGAGGCGTTCGCGAGGAGCTCAAGCACTTGAAGGCAAAATCGAAGAGCAC 840
841 AACCACTACAGAGCAGCTGAGATTTGCGCAGAGAGCTGAGGCAAGCGAGAGCGTG 900
841 AACCACTACAGAGCAGCTGAGATTTGCGCAGAGAGCTGAGGCAAGCGAGAGCGTG 900
901 GCGAGCGGAGGCTGTGAGCGCGCAGCGCGAGAGCAGCGCTGCTGAGGAGGCGGAGCC 960
901 GCGAGCGGAGGCTGTGAGCGCGCAGCGCGAGAGCAGCGCTGCTGAGGAGGCGGAGCC 960
961 AAGGAGCTGAGGCTTACAGGCTGAAGAGCATCTGAGGAGCTGTCGAGGAGTCTCCAGA 1020
961 AAGGAGCTGAGGCTTACAGGCTGAAGAGCATCTGAGGAGCTGTCGAGGAGTCTCCAGA 1020
1021 GCTGCGCGAGGAGCAAAACCGGTGCAACCCGCGCTCCAGAGCATGTGCCCAACAG 1080
1021 GCTGCGCGAGGAGCAAAACCGGTGCAACCCGCGCTCCAGAGCATGTGCCCAACAG 1080
1081 TGGAGCTCCCCCCCCAAGCGCTTTCGATTTGGCCCCCTGACAAGGCCATCAACAGGAA 1140
1081 TGGAGCTCCCCCCCCAAGCGCTTTCGATTTGGCCCCCTGACAAGGCCATCAACAGGAA 1140
1141 CAGTGCAGGCGCGCGCTGCTGCTACATCCCTGTCAAAAGCAGGGCTGAGGGAGCCAG 1200
1141 CAGTGCAGGCGCGCGCTGCTGCTACATCCCTGTCAAAAGCAGGGCTGAGGGAGCCAG 1200
1201 ATGGGCGAGGCTGTGCTTCTCCACCGAGCTACCCAGCTPACAGCTGAGAGACCTG 1260

1201 ATGGGCGAGGCTGTGCTTCTTCCACCCAGCTACCCAGCTPACAGCTGAGAGACCTG 1260
1261 AGCTCTCTGAAATGGGCTACACGCGCACTGACCCGTAACCCGTAACCCGTAACCTTCTTCCCC 1320
1261 AGCTCTCTGAAATGGGCTACACGCGCACTGACCCGTAACCCGTAACCCGTAACCTTCTTCCCC 1320
1321 AAGGACATCTCTGACCTCGGCTGAGACGTGATGATGGAGACTGAGAAACCGCTCCACTTC 1380
1321 AAGGACATCTCTGACCTCGGCTGAGACGTGATGATGGAGACTGAGAAACCGCTCCACTTC 1380
1381 ACGATCAAGATCCAGCTAACAGGGCGCTACAGGCTGCGCTTGGAGACCCGCGGTGTCAC 1440
1381 ACGATCAAGATCCAGCTAACAGGGCGCTACAGGCTGCGCTTGGAGACCCGCGGTGTCAC 1440
1441 AGCCGGGACCGTCCCTCCACTCTACAGCGTGGAGTTCTCCGAGAGCGCTTCGGGGTGATC 1500
1441 AGCCGGGACCGTCCCTCCACTCTACAGCGTGGAGTTCTCCGAGAGCGCTTCGGGGTGATC 1500
1501 GTGCAACCGGACGTGGACGGCCGTGTCTGTAACAGACGGTGGCGCCCTGTCTTCTT 1560
1501 GTGCAACCGGACGTGGACGGCCGTGTCTGTAACAGACGGTGGCGCCCTGTCTTCTT 1560
1561 GCGGACAGTTCCTTCAGCTGTCCAGCTGCTGCGCTGCGCTGCGAGTATACAGGCTCGCC 1620
1561 GCGGACAGTTCCTTCAGCTGTCCAGCTGCTGCGCTGCGCTGCGAGTATATCACAGGCTCGCC 1620
1621 GAGCACTCAGTCCCTGATGCTCAGACCACTGAGACCGAGTACACCTGTGGAAACCGG 1680
1621 GAGCACTCAGTCCCTGATGCTCAGACCACTGAGACCGAGTACACCTGTGGAAACCGG 1680
1681 GACCTTGGCGCCACCGCCGGTGGAACTCTACGGGTCTACCCCTTCTACCTGGCGCTG 1740
1681 GACCTTGGCGCCACCGCCGGTGGAACTCTACGGGTCTACCCCTTCTACCTGGCGCTG 1740
1741 GAGGACGGGCTGGCGCACCGGGTGTCTCTGCTAAGCAAGCAATGCCATGATGTGTC 1800
1741 GAGGACGGGCTGGCGCACCGGGTGTCTCTGCTAAGCAAGCAATGCCATGATGTGTC 1800
1801 CTGACGCGAGCGCTGCCCTTAGCTGAGGCTGCAAGGTGGGATCTGATGATGTCTACATC 1860
1801 CTGACGCGAGCGCTGCCCTTAGCTGAGGCTGCAAGGTGGGATCTGATGATGTCTACATC 1860
1861 TTCCTGGGCGGCGCAAGAGCGTGTGTCAGCAGTACCTTGGAGTGTGGGATACCGG 1920
1861 TTCCTGGGCGGCGCAAGAGCGTGTGTCAGCAGTACCTTGGAGTGTGGGATACCGG 1920
1921 TTCATGCGGCGCATCTGGGCGCTGGGCTTCCACTGTCGCGTGGGCTTCTCTCCACC 1980
1921 TTCATGCGGCGCATCTGGGCGCTGGGCTTCCACTGTCGCGTGGGCTTCTCTCCACC 1980
1981 GCTATCACCGCGCAGGTGGTGGAGAACTAGACAGGGCGGCTTCCCCCTGGACGTCCTCAA 2040
1981 GCTATCACCGCGCAGGTGGTGGAGAACTAGACAGGGCGGCTTCCCCCTGGACGTCCTCAA 2040
2041 TGGAAACGACCTGGACTCATGGAATCCCGGAGGACTTCACTTCAAAAGAGTGGCTTC 2100
2041 TGGAAACGACCTGGACTCATGGAATCCCGGAGGACTTCACTTCAAAAGAGTGGCTTC 2100
2101 CCGGACTTCCCGGCGCATGCTGAGAGTGTGACAGGGCGGCGCTTACATGATGATC 2160
2101 CCGGACTTCCCGGCGCATGCTGAGAGTGTGACAGGGCGGCGCTTACATGATGATC 2160
2161 GTGGATCTTGCATCAGACGCTGCGGCGCTGCGCGGAGCTACAGGCGCTTACGACGAGGT 2220
2161 GTGGATCTTGCATCAGACGCTGCGGCGCTGCGCGGAGCTACAGGCGCTTACGACGAGGT 2220
2221 CTGCGGAGGGGGTTTTTTCATCAACGAGACCGGCGAGCTGATTTGGGAAGGTATGG 2280
2221 CTGCGGAGGGGGTTTTTTCATCAACGAGACCGGCGAGCTGATTTGGGAAGGTATGG 2280
2281 CCGGGGTCACCTGCTTCCCGACTTCCCGACTTCCCGACTTCCCGACTTCCCGACTTCCCG 2340
2281 CCGGGGTCACCTGCTTCCCGACTTCCCGACTTCCCGACTTCCCGACTTCCCGACTTCCCG 2340

|||||
361 CAGGTGACAGCAACTCCCTCAGTGGCACCCAGGAAGACGGGTGGATGACCCAGGCTG 420
QY
421 GAAGAAGCTGTGGCAGCAAGCGGAGACCTCTGGGAATTTCTCGGCGGAAGAACTGGACAG 480
DB
421 GAAGAAGCTGTGGCAGCAAGCGGAGACCTCTGGGAATTTCTCGGCGGAAGAACTGGACAG 480
QY
481 CTCTGGCGGGAGTTCTCTGCATCACAAAGAGAAGTTTCAGAGTACAAAGCTCTGTCTGGAG 540
DB
481 CTCTGGCGGGAGTTCTCTGCATCACAAAGAGAAGTTTCAGAGTACAAAGCTCTGTCTGGAG 540
QY
541 ACCCTGAGCAGGACCGAAGAAATCCACGAGAACGTCATTAGCCCTCGGACCTGAGCGAC 600
DB
541 ACCCTGAGCAGGACCGAAGAAATCCACGAGAACGTCATTAGCCCTCGGACCTGAGCGAC 600
QY
601 ATCAAGGCGAGGCTCTGTGCACAGCAGGACACGAGAGTCAAGGAGAGCTGGCGAGCATC 660
DB
601 ATCAAGGCGAGGCTCTGTGCACAGCAGGACACGAGAGTCAAGGAGAGCTGGCGAGCATC 660
QY
661 AACCAAGGCGCTGTGACCGGCTGTGGAGGCTCAGCCACGAGGCTACAGCACTGAGGCTGAG 720
DB
661 AACCAAGGCGCTGTGACCGGCTGTGGAGGCTCAGCCACGAGGCTACAGCACTGAGGCTGAG 720
QY
721 TTCTGAGGAGCCAGAGGTGATTGACTGTGGGACTTGGGAGCTTGGGAGCTTGGGAGCTTGGGAG 780
DB
721 TTCTGAGGAGCCAGAGGTGATTGACTGTGGGACTTGGGAGCTTGGGAGCTTGGGAGCTTGGGAG 780
QY
781 AAGGAGCTGGAGGCTTCCGGAGGAGCTCAAGCACTTCGAGGCCAAATCAGAGAGCAC 840
DB
781 AAGGAGCTGGAGGCTTCCGGAGGAGCTCAAGCACTTCGAGGCCAAATCAGAGAGCAC 840
QY
841 AACCACTACAGAGCAGCTGGAGATTGGGCGACGAGAGCTGAGGCACTGAGGAGGCGTG 900
DB
841 AACCACTACAGAGCAGCTGGAGATTGGGCGACGAGAGCTGAGGCACTGAGGAGGCGTG 900
QY
901 GGCGAGGCGAGGCTGTGAGCCGAGCAGCCGAGAGAGCAAGCAAGCTGTGGAGGGGCGGACC 960
DB
901 GGCGAGGCGAGGCTGTGAGCCGAGCAGCCGAGAGAGCAAGCAAGCTGTGGAGGGGCGGACC 960
QY
961 AAGGAGCTGGGCTACACGCTGAGAGAGCATCTGAGGAGCTTCCGAGGAGTCTCCAGA 1020
DB
961 AAGGAGCTGGGCTACACGCTGAGAGAGCATCTGAGGAGCTTCCGAGGAGTCTCCAGA 1020
QY
1021 GCTCGCGCGAGGAGCAAAACCGGTGCAACCCGCGGCTCCAGAGAGCTGCCACACAG 1080
DB
1021 GCTCGCGCGAGGAGCAAAACCGGTGCAACCCGCGGCTCCAGAGAGCTGCCACACAG 1080
QY
1081 TGGAGCTCCCCCAGAGCCGCTTTCGATTGGGCGGCTGACAGAGCCATCACCCAGGAA 1140
DB
1081 TGGAGCTCCCCCAGAGCCGCTTTCGATTGGGCGGCTGACAGAGCCATCACCCAGGAA 1140
QY
1141 CAGTGGAGGCGCGGCTGTCTGCTACATCCCTGCAAGAGAGGAGGCTGAGGAGGAGGAG 1200
DB
1141 CAGTGGAGGCGCGGCTGTCTGCTACATCCCTGCAAGAGAGGAGGCTGAGGAGGAGGAG 1200
QY
1201 ATGGGAGAGCCCTGTGTCTTCTTCCACCCAGCTACCCAGCTACAAAGCTGGAGAACCTG 1260
DB
1201 ATGGGAGAGCCCTGTGTCTTCTTCCACCCAGCTACCCAGCTACAAAGCTGGAGAACCTG 1260
QY
1261 AGCTCTCTGAAATGGGCTACAGGCGCCCTGACCCGCTGACCCGCTGACCCGCTTCTTCCC 1320
DB
1261 AGCTCTCTGAAATGGGCTACAGGCGCCCTGACCCGCTGACCCGCTGACCCGCTTCTTCCC 1320
QY
1321 AAGGAGATCCTGACCTGCGGCTGAGAGCTGATGAGAGCTGAGAGCTGAGAGCTGAGAGCTG 1380
DB
1321 AAGGAGATCCTGACCTGCGGCTGAGAGCTGATGAGAGCTGAGAGCTGAGAGCTGAGAGCTG 1380
QY
1381 ACGATCAAAAGATCCAGCTAACAGGCGCTACGAGGTGCCCTTGGAGACCCCGGCTGTCCAC 1440
DB
1381 ACGATCAAAAGATCCAGCTAACAGGCGCTACGAGGTGCCCTTGGAGACCCCGGCTGTCCAC 1440
QY
1441 AGCGGGGACCGTCCCACTTACAGCTGAGAGTCTCCGAGGAGCCCTTCCGGGTGATC 1500
|||||

1441 AGCGGGGACCGTCCCACTCTTACAGCGTGGAGTTCTCCGAGGAGCCCTTCGGGGTGATC 1500
QY
1501 GTGACCCGCGAGCTGAGCGGCGCTGTCTGTGAACAGAGCGTGGCGGCCCTGTCTTCTT 1560
DB
1501 GTGACCCGCGAGCTGAGCGGCGCTGTCTGTGAACAGAGCGTGGCGGCCCTGTCTTCTT 1560
QY
1561 GCGGACCAAGTTCTTTCAGCTGTCCACTCGCTGCCCTCGCAGTATATCACAGGCTCGCC 1620
DB
1561 GCGGACCAAGTTCTTTCAGCTGTCCACTCGCTGCCCTCGCAGTATATCACAGGCTCGCC 1620
QY
1621 GAGCACTCTAGTCTCCTGATGCTAGCACCAGCTGGAACAGGATCACCTGTGGAAACCG 1680
DB
1621 GAGCACTCTAGTCTCCTGATGCTAGCACCAGCTGGAACAGGATCACCTGTGGAAACCG 1680
QY
1681 GACCTTGGCGCCAGCCCGCTGCGAACTCTACGGGTCTCACCTTTCTACCTGGCGCTG 1740
DB
1681 GACCTTGGCGCCAGCCCGCTGCGAACTCTACGGGTCTCACCTTTCTACCTGGCGCTG 1740
QY
1741 GAGGACGCGGCTCGGCAACGCGGCTTCTCTGTAAACAGCAATGCCATGATGTGTC 1800
DB
1741 GAGGACGCGGCTCGGCAACGCGGCTTCTCTGTAAACAGCAATGCCATGATGTGTC 1800
QY
1801 CTGACCGGAGCCCTGCGCTTAGCTGGAGTCCAGAGTGGATCCTGGATCTTACATC 1860
DB
1801 CTGACCGGAGCCCTGCGCTTAGCTGGAGTCCAGAGTGGATCCTGGATCTTACATC 1860
QY
1861 TTCTTGGGCGCCAGAGCCCAAGAGCGTGTGAGAGTACCTGAGAGTGTGGATACCG 1920
DB
1861 TTCTTGGGCGCCAGAGCCCAAGAGCGTGTGAGAGTACCTGAGAGTGTGGATACCG 1920
QY
1921 TTCTATGCGCCATACTGCGGCTTCCACTGTGCGCTGGGCTTCTCTCCACC 1980
DB
1921 TTCTATGCGCCATACTGCGGCTTCCACTGTGCGCTGGGCTTCTCTCCACC 1980
QY
1981 GCTATCAACCGGAGGCTGTGGAGAACATGACAGGAGGCTTCCCTTCCAGCTCCAA 2040
DB
1981 GCTATCAACCGGAGGCTGTGGAGAACATGACAGGAGGCTTCCCTTCCAGCTCCAA 2040
QY
2041 TGGAAACGAGCTGGACTACATGAGACTCCCGAGGAGCTTCAAGTCAACAGAGTGGCTC 2100
DB
2041 TGGAAACGAGCTGGACTACATGAGACTCCCGAGGAGCTTCAAGTCAACAGAGTGGCTC 2100
QY
2101 GCGGACTTCCCGGCTCATGAGAGTGCACAGGCGGCGGCTTACATGATGATC 2160
DB
2101 GCGGACTTCCCGGCTCATGAGAGTGCACAGGCGGCGGCTTACATGATGATC 2160
QY
2161 GTGGATCTGCTCAGCAGCTCGGCGCTCGCGGAGCTACAGGCTTACAGGAGGCT 2220
DB
2161 GTGGATCTGCTCAGCAGCTCGGCGCTCGCGGAGCTACAGGCTTACAGGAGGCT 2220
QY
2221 CTGCGGAGGCGGCTTTCATCACAAACGAGAGCCGCGCTGATTTGGGAGGATG 2280
DB
2221 CTGCGGAGGCGGCTTTCATCACAAACGAGAGCCGCGCTGATTTGGGAGGATG 2280
QY
2281 CCGGCTCCAGCTGCTTCCCGACTTCCAAACCCACAGGCTTGGGAGGAG 2340
DB
2281 CCGGCTCCAGCTGCTTCCCGACTTCCAAACCCACAGGCTTGGGAGGAG 2340
QY
2341 ATGGTGGCTGAGTTCATGACAGGCTTTCGAGCGCTTGGATGATGATGATGATG 2400
DB
2341 ATGGTGGCTGAGTTCATGACAGGCTTTCGAGCGCTTGGATGATGATGATGATG 2400
QY
2401 CTTTCAAACTTTCATCAGAGGCTTTCGAGCGGCTTCCCAACAAATGAGCTGGAGAAC 2460
DB
2401 CTTTCAAACTTTCATCAGAGGCTTTCGAGCGGCTTCCCAACAAATGAGCTGGAGAAC 2460
QY
2461 CCCTAGCTGCTTGGGCTTGGGAGGAGCCCTTCCAGGCGGCGGCTTGGGCTTCCAG 2520
DB
2461 CCCTAGCTGCTTGGGCTTGGGAGGAGCCCTTCCAGGCGGCGGCTTGGGCTTCCAG 2520
QY
2521 CACCAAGTTCCTTCCACACACTACAACTTACAGGCTTACAGGCTTACAGGCTTAC 2580
DB
2521 CACCAAGTTCCTTCCACACACTACAACTTACAGGCTTACAGGCTTACAGGCTTAC 2580

Query Match	71.5%	Score 2647.8	DB 18	Length 3846
Best Local Similarity	99.7%	Pred. No. 0		
Matches 2652	Conservative 0	Mismatches 7	Indels 0	Gaps 0
Qy	1036	GAACCGCGTGACACACCCCGCGCTCCACAGACAGTGCACACAGTGCAGCTGCCCGCC	1095	
Db	640	GATGCCCAGGCACACCCCGCGCTCCACAGCAGTGCACACAGTGGAGCTCCCCCCC	699	
Qy	1096	AACAGCGCTTCGATTCGGCCCTTGAACAAGGCCATCAACCAGGAACAAGTGGAGGCCCGC	1155	
Db	700	AACAGCGCTTCGATTCGGCCCTTGAACAAGGCCATCAACCAGGAACAAGTGGAGGCCCGC	759	
Qy	1156	GGCTGCTGTACATCCCTGCAAAAGCAGGGCTGCAGGAGGCCACAGATGGGGCAGCCCTGG	1215	
Db	760	GGCTGCTGTACATCCCTGCAAAAGCAGGGCTGCAGGAGGCCACAGATGGGGCAGCCCTGG	819	
Qy	1216	TGCTTCTTCCACCCAGCTACCCAGCTACAAGCTGGAGAACCTTGAGCTCCTCTGAAATG	1275	
Db	820	TGCTTCTTCCACCCAGCTACCCAGCTACAAGCTGGAGAACCTTGAGCTCCTCTGAAATG	879	
Qy	1276	GGCTACAAGGCCACCTGACCCGTACACCCCCACCTTCTTCCCCAGGACATCCTCGACC	1335	
Db	880	GGCTACAAGGCCACCTGACCCGTACACCCCCACCTTCTTCCCCAGGACATCCTCGACC	939	
Qy	1336	CTGGCGCTGACAGCTGATGTAGAGACTGAGAACCGCCTCCACTTTCACGATCAAAAGATCCA	1395	
Db	940	CTGGCGCTGACAGCTGATGTAGAGACTGAGAACCGCCTCCACTTTCACGATCAAAAGATCCA	999	
Qy	1396	GCTAACAGGCGCTACGAGGTGCCCTTGGAGAACCCCGGTGTGCACAGCGGGGACCGGTCC	1455	
Db	1000	GCTAACAGGCGCTACGAGGTGCCCTTGGAGAACCCCGGTGTGCACAGCGGGGACCGGTCC	1059	
Qy	1456	CCACTCTACAGCGTGGAGTTCTCCGAGAGGCCCTTCGGGGGTGATCGTGCAACCGGACGTG	1515	
Db	1060	CCACTCTACAGCGTGGAGTTCTCCGAGAGGCCCTTCGGGGGTGATCGTGCAACCGGACGTG	1119	
Qy	1516	GACGGCGCGTGTGCTGTAACAACAGCAGGTGGCGCCCTTGTTCTTTTTCGGACACAGTTCCCT	1575	
Db	1120	GACGGCGCGTGTGCTGTAACAACAGCAGGTGGCGCCCTTGTTCTTTTTCGGACACAGTTCCCT	1179	
Qy	1576	CAGCTGTCCACCTCGCTGCCCTCCGATATATACAGGCCCTCGCGGACACCTCAGTCCC	1635	
Db	1180	CAGCTGTCCACCTCGCTGCCCTCCGATATATACAGGCCCTCGCGGACACCTCAGTCCC	1239	
Qy	1636	CTGATGTCTCAGCACAGCTGACACAGGATCACCTCTGGAAACCGGGACCTTGGCCCGACG	1695	
Db	1240	CTGATGTCTCAGCACAGCTGACACAGGATCACCTCTGGAAACCGGGACCTTGGCCCGACG	1299	
Qy	1696	CCCGGTGCGAAACCTTACGGGTCTCAACCTTTTCTACTGGCGCTTGGAGGACGGCGGTGCG	1755	
Db	1300	CCCGGTGCGAAACCTTACGGGTCTCAACCTTTTCTACTGGCGCTTGGAGGACGGCGGTGCG	1359	
Qy	1756	GCACACGGGGTTCCTGCTAAACAGCAATGCCATGGATGGTTCCTGCAGCCGACGACCT	1815	
Db	1360	GCAACAGGGGTTCCTGCTAAACAGCAATGCCATGGATGGTTCCTGCAGCCGACGACCT	1419	
Qy	1816	GCCCTTAGCTGGAGGTGCACAGGTGGGATCCTGGATGCTACATCTTCTGGGCGCCAGAG	1875	

```
Db 2500 TTCAGGAGCGCGCCAGCAGGCGCATGAGAGGCGCCCTCACCTGCGCTACGCACTCCTC 2559
QY CCCCACTCTACACACTGTTCCACACAGGCCACGTCGCGGGGAGACCGTGGCGCGGCC 3015
Db |||||
Db 2560 CCCCACTCTACACACTGTTCCACACAGGCCACGTCGCGGGGAGACCGTGGCGCGGCC 2619
QY CTCCTTCTGGAGTTCCCAAGGACTTAGCACCCTGAGCTGTGGAACCAAGCTCTCTGTGG 3075
Db CTCCTTCTGGAGTTCCCAAGGACTTAGCACCCTGAGCTGTGGAACCAAGCTCTCTGTGG 2679
QY GGGGAGCGCTGCTCATACCCAGTCTCCAGGCGGGAAGCGCGAAGTGAAGTGAAGTGA 3135
Db GGGGAGCGCTGCTCATACCCAGTCTCCAGGCGGGAAGCGCGAAGTGAAGTGAAGTGA 2739
QY TTCCCTCTGGGACATAGTACGACCTGACAGCGGTGCCAATAGAGGCCCTTGGCAGCCTC 3195
Db TTCCCTCTGGGACATAGTACGACCTGACAGCGGTGCCAATAGAGGCCCTTGGCAGCCTC 2799
QY CACCCCCACCTGACGTCCTCCGTGAGCGAGCCATCCACGAGGAGGGGAGTGGGTGACG 3255
Db CACCCCCACCTGACGTCCTCCGTGAGCGAGCCATCCACGAGGAGGGGAGTGGGTGACG 2859
QY CTGCGGCGCCCTTGACACCATCAACGTCACCTCCGCGCTGGGTACATCATCCGCCCTG 3315
Db CTGCGGCGCCCTTGACACCATCAACGTCACCTCCGCGCTGGGTACATCATCCGCCCTG 2919
QY CAGGCGCTGGCTCAACACACAGAGTCCCGCAGAGCCCATGAGCCCTGGCTGTGGCC 3375
Db CAGGCGCTGGCTCAACACACAGAGTCCCGCAGAGCCCATGAGCCCTGGCTGTGGCC 2979
QY CTAAACCAAGGTGGAGAGCGCGAGGGAGTGTCTTGGGACGATGGAGAGCCTGGAA 3435
Db CTGACCAAGGTGGAGAGCGCGAGGGAGTGTCTTGGGACGATGGAGAGCCTGGAA 3039
QY GTGCTGGAGGAGGGGCTTACACAGGTCACTTCTCGCCAGGAAATACACGATCGTG 3495
Db GTGCTGGAGGAGGGGCTTACACAGGTCACTTCTCGCCAGGAAATACACGATCGTG 3099
QY AATGAGCTGTAGTGTGACCAAGTGGAGGAGCTGGGCTGCAGCTGCAGAGGTGACTGTC 3555
Db AATGAGCTGTAGTGTGACCAAGTGGAGGAGCTGGGCTGCAGCTGCAGAGGTGACTGTC 3159
QY CTGGGCGTGGCCACGCGGCCAGCAGGTCTCTCCAAAGGTGTCCTGTCTCCAACTTC 3615
Db CTGGGCGTGGCCACGCGGCCAGCAGGTCTCTCTCCAAAGGTGTCCTGTCTCCAACTTC 3219
QY ACCTAGACCCCGACACCAAGTCTCGGACATCTGTCTCGCTGTGATGGAGAGCAG 3675
Db ACCTAGACCCCGACACCAAGTCTCGGACATCTGTCTCGCTGTGATGGAGAGCAG 3279
QY TTCTCTGTCAGCTGGTGT 3694
Db TTCTCTGTCAGCTGGTGT 3298
```

RESULT 5

```
US-10-874-242-38
; Sequence 38, Application US/10874242
; Publication No. US20050026252A1
; GENERAL INFORMATION:
; APPLICANT: ESTES, SCOTT
; APPLICANT: ZHANG, WEIQUN
; TITLE OF INVENTION: NOVEL PROMOTERS AND USES THEREOF
; FILE REFERENCE: 07680.0027-00000
; CURRENT APPLICATION NUMBER: US/10/874,242
; PRIOR FILING DATE: 2004-06-24
; PRIOR APPLICATION NUMBER: 60/480,768
; PRIOR FILING DATE: 2003-06-24
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 38
; LENGTH: 2859
; TYPE: DNA
```

```
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: cdna encoding alpha-glucosidase.
US-10-874-242-38
```

Query Match 71.4%; Score 2643.4; DB 19; Length 2859;

Best Local Similarity 99.6%; Pred. No. 0; Mismatches 11; Indels 0; Gaps 0;

Matches 2650; Conservative 0;

```
QY 1036 GAAACCGGTGACACCCCGGCGCTCCAGAGCAGTGGCCACACAGTGCAGCTGCCCCCC 1095
Db |||||
QY 199 GATGCCAGGACACCCCGGCGCTCCAGAGCAGTGGCCACACAGTGCAGCTGCCCCCC 258
Db |||||
QY 1096 AACAGCCGCTTCGATTGCGCCCTTGACAAAGGCGCATCACCCAGGAAACAGTGCAGGCGCCG 1155
Db AACAGCCGCTTCGATTGCGCCCTTGACAAAGGCGCATCACCCAGGAAACAGTGCAGGCGCCG 318
QY 1156 GGCTGCTGTATACCTCTGCAAAAGCAGGGGCTGCAAGGAGGCCAGATGGGGGAGCGCTGG 1215
Db |||||
QY 319 GGCTGCTGTATACCTCTGCAAAAGCAGGGGCTGCAAGGAGGCCAGATGGGGGAGCGCTGG 378
QY 1216 TGCTTTTCCCAACCCAGCTTACCCAGCTACAAAGCTGGAGAACCTGAGCTCTCTGAAATG 1275
Db TGCTTTTCCCAACCCAGCTTACCCAGCTACAAAGCTGGAGAACCTGAGCTCTCTGAAATG 438
QY 1276 GGCTACAGCGGCACCCCTGACCCGCTACCCACCCCACTTCTTCCCAAGGACATCTCTGACC 1335
Db GGCTACAGCGGCACCCCTGACCCGCTACCCACCCCACTTCTTCCCAAGGACATCTCTGACC 498
QY 1336 CTGCGGCTGGAGCTGATATGAGAGCTGAGAAACCGCTCCACATTCACGATCAAGATCCA 1395
Db CTGCGGCTGGAGCTGATATGAGAGCTGAGAAACCGCTCCACATTCACGATCAAGATCCA 558
QY 1396 GCTAACAGCGCTACGAGTGCCCTTGGAGACCCCGCTGTCACAGCGGACACCGTCC 1455
Db GCTAACAGCGCTACGAGTGCCCTTGGAGACCCCGCTGTCACAGCGGACACCGTCC 618
QY 1456 CCACCTTACAGCGTGGAGTTCTCGAGGAGCCCTTTCGGGGTGATCGTCACCGGAGCTG 1515
Db CCACCTTACAGCGTGGAGTTCTCTGAGGAGCCCTTTCGGGGTGATCGTCACCGGAGCTG 678
QY 1516 GACGCGCGGCTGCTGTAACACGACGCGTGGCGGCCCTGTTCTTTGCGGACAGTTCCTT 1575
Db GACGCGCGGCTGCTGTAACACGACGCGTGGCGGCCCTGTTCTTTGCGGACAGTTCCTT 738
QY 1576 CAGCTGTCCACTGCTGCGCTGCGAGTATATACAGGCTCGCGAGCACCTCAGTCC 1635
Db CAGCTGTCCACTGCTGCGCTGCGAGTATATACAGGCTCGCGAGCACCTCAGTCC 798
QY 1636 CTGATGCTCAGCACCCAGCTGGACAGGATCACCCCTGTGGAACCCGGGACCTTGCGCCACG 1695
Db CTGATGCTCAGCACCCAGCTGGACAGGATCACCCCTGTGGAACCCGGGACCTTGCGCCACG 858
QY 1696 CCGGTGCGAACCTCTACGGGTCTACCCCTTTCATCTGCGCTGCGGAGACCGCGGCTG 1755
Db CCGGTGCGAACCTCTACGGGTCTACCCCTTTCATCTGCGCTGCGGAGACCGCGGCTG 918
QY 1756 GCACACGGGGTGTCTCTGCTAAACAGCAATGCTATGATGTGCTGCTGAGCCGAGCCCT 1815
Db GCACACGGGGTGTCTCTGCTAAACAGCAATGCTATGATGTGCTGCTGAGCCGAGCCCT 978
QY 1816 GCGCTTAGCTGGAGGTGACAGTGGGATCTCTGATGTCTACATCTTCTCGGGCCAGAG 1875
Db GCGCTTAGCTGGAGGTGACAGTGGGATCTCTGATGTCTACATCTTCTCGGGCCAGAG 1038
QY 1876 CCCAAGAGCGTGGTGCAGAGTACTGAGACGTTGTGGGATACCCGTTATGCGGCGCATAC 1935
Db CCCAAGAGCGTGGTGCAGAGTACTGAGACGTTGTGGGATACCCGTTATGCGGCGCATAC 1098
QY 1936 TGGGGCTGGGCTTCCACTGTGCGGCTGGGGCTACTCTCCACCGCTATACCCGCCAG 1995
Db TGGGGCTGGGCTTCCACTGTGCGGCTGGGGCTACTCTCTCCACCGCTATACCCGCCAG 1158
```

```
QY 1996 GTGGTGGAGAACATGACACAGGGGCCCACTTCCCTCTGGAAGCTCCAAATGGAAACGACTTGGAC 2055
Db 1159 GTGGTGGAGAACATGACACAGGGGCCCACTTCCCTCTGGAAGCTCCAAATGGAAACGACTTGGAC 1218
QY 2056 TACATGAGACTCCCGGAGGACTTACGTTTCAACAAGGATGGCTTCCGGGACTTCCCGGCC 2115
Db 1219 TACATGAGACTCCCGGAGGACTTACGTTTCAACAAGGATGGCTTCCGGGACTTCCCGGCC 1278
QY 2116 ATGGTGCAGGAGTGCACCAAGGCGCGCGCTTACATGATGATCGTGGATCTCTGCCATC 2175
Db 1279 ATGGTGCAGGAGTGCACCAAGGCGCGCGCTTACATGATGATCGTGGATCTCTGCCATC 1338
QY 2176 AGCAGCTCGGGCCCTCGCGGGAGCTACAGGCCCTCAAGCAGAGGTCTCGGAGGGGGTT 2235
Db 1339 AGCAGCTCGGGCCCTCGCGGGAGCTACAGGCCCTCAAGCAGAGGTCTCGGAGGGGGTT 1398
QY 2236 TTCAATCAACAAAGAGACCGGCGAGCGCTGATTTGGGAAGGTAAGGCCCGGCTCACTGCG 2295
Db 1399 TTCAATCAACAAAGAGACCGGCGAGCGCTGATTTGGGAAGGTAAGGCCCGGCTCACTGCG 1458
QY 2296 TTCCCGGACTTCAACCAACCCACAGCCCTGGCTGGTGGAGGACATGGTGGCTGAGTTT 2355
Db 1459 TTCCCGGACTTCAACCAACCCACAGCCCTGGCTGGTGGAGGACATGGTGGCTGAGTTT 1518
QY 2356 CATGACAGGTGCTTCCAGCGGCTTGTGATTTGACATGAAAGAGCCCTTCCAACTTCATC 2415
Db 1519 CATGACAGGTGCTTCCAGCGGCTTGTGATTTGACATGAAAGAGCCCTTCCAACTTCATC 1578
QY 2416 AGAGGCTCTGAGGACGCGTCCCAACAAATAGCTGGAGAACCCCTAGCTGCTGGG 2475
Db 1579 AGGGGCTCTGAGGACGCGTCCCAACAAATAGCTGGAGAACCCCTAGCTGCTGGG 1638
QY 2476 GTGGTTGGGGGACCCCTCCAGGGGCGCACCATCTGTGCTCTCCAGCCACAGATTTCTCTCC 2535
Db 1639 GTGGTTGGGGGACCCCTCCAGGGGCGCACCATCTGTGCTCTCCAGCCACAGATTTCTCTCC 1698
QY 2536 ACACACTACAACTCTGACAACTCTACGCGCTGACCGAAGCCATCGCTCCCAAGGGG 2595
Db 1699 ACACACTACAACTCTGACAACTCTACGCGCTGACCGAAGCCATCGCTCCCAAGGGG 1758
QY 2596 CTGGTGAAGGCTCGGGGACACCCCATTTGTGATCTCCGCTCGACCTTTGCTGGCCAC 2655
Db 1759 CTGGTGAAGGCTCGGGGACACCCCATTTGTGATCTCCGCTCGACCTTTGCTGGCCAC 1818
QY 2656 GGCCGATAGCCGCGCACTGGACGGGGAGCTGTGGAGTCTCTGGGAGCAGCTCGGCTCC 2715
Db 1819 GGCCGATAGCCGCGCACTGGACGGGGAGCTGTGGAGTCTCTGGGAGCAGCTCGGCTCC 1878
QY 2716 TCCGTGCCAGAAATCTCGCAGTTTAACTGCTGGGGTGTCTGTGCTGGGGCGGACGTC 2775
Db 1879 TCCGTGCCAGAAATCTCGCAGTTTAACTGCTGGGGTGTCTGTGCTGGGGCGGACGTC 1938
QY 2776 TCGGGCTTCTGGGCAACCTCAGAGGAGCTGTGTGCTGGGCTGGACCCAGCTGGGGGCC 2835
Db 1939 TCGGGCTTCTGGGCAACCTCAGAGGAGCTGTGTGCTGGGCTGGACCCAGCTGGGGGCC 1998
QY 2836 TTCTACCCCTTCAATGGGGAACCAACAGCTCTCTCAGTCTGCGCCAGGAGCGGTACAGC 2895
Db 1999 TTCTACCCCTTCAATGGGGAACCAACAGCTCTCTCAGTCTGCGCCAGGAGCGGTACAGC 2058
QY 2896 TTCAGGAGCGGGCCAGCAGGCGCATGAGGAAGGCCCTCACCCTGGCTAGCGACTCTCTC 2955
Db 2059 TTCAGGAGCGGGCCAGCAGGCGCATGAGGAAGGCCCTCACCCTGGCTAGCGACTCTCTC 2118
QY 2956 CCCCACCTTACACATGTTTCCACGAGGCCAGCTCGCGGGGAGACCGTGGCCGGGCC 3015
Db 2119 CCCCACCTTACACATGTTTCCACGAGGCCAGCTCGCGGGGAGACCGTGGCCGGGCC 2178
QY 3016 CTCTTCTGGAGTTTCCCAAGGACTCTAGCAGCTGAGCTGTGGAACCAACAGCTCTCTGTG 3075
Db 2179 CTCTTCTGGAGTTTCCCAAGGACTCTAGCAGCTGAGCTGTGGAACCAACAGCTCTCTGTG 2238
QY 3076 GGGAGGGCCCTGCTCATATCAACCCAGTGTCTCCAGGCCGGGAAGGCCGGAAGTGAATGCGCTAC 3135
```

```
Db 2239 GGGGAGGCCCTGCTCATATCAACCCAGTGTCTCAGGCGGGGAAGGCCGGAAGTGAATGCGCTAC 2298
QY 3136 TTCCCTTGGGCAACATGTTACGCTGACAGCGTGCCTCAATAGAGCCCTTTGGCAGCTC 3195
Db 2299 TTCCCTTGGGCAACATGTTACGCTGACAGCGTGCCTCAATAGAGCCCTTTGGCAGCTC 2358
QY 3196 CCAACCCCACTGCACTCCCGTGGAGCCAGGCATCCACAGCAGGGGCGAGTGGGTGACG 3255
Db 2359 CCAACCCCACTGCACTCCCGTGGAGCCAGGCATCCACAGCAGGGGCGAGTGGGTGACG 2418
QY 3256 CTGCGGGCCCTTGGAGCAGCATCAAGCTCCACCTCCGGGCTGGGTACATCATCCCCCTG 3315
Db 2419 CTGCGGGCCCTTGGAGCAGCATCAAGCTCCACCTCCGGGCTGGGTACATCATCCCCCTG 2478
QY 3316 CAGGGCCCTGGCTTCAACCAACAGAGTCCCGCAGCAGCCATGCGCTGTGGCC 3375
Db 2479 CAGGGCCCTGGCTTCAACCAACAGAGTCCCGCAGCAGCCATGCGCTGTGGCC 2538
QY 3376 CTAAACAAAGGTGGAGAGGCCCGAGGGAGCTGTTCTGGGACGATGGAGAGCCTGGAA 3435
Db 2539 CTGACCAAGGTGGAGAGGCCCGAGGGAGCTGTTCTGGGACGATGGAGAGCCTGGAA 2598
QY 3436 GTGCTGGAGCGAGGGCCCTTACACAGGTCTCTTCTGCGCCAGGAATACACGATCGTG 3495
Db 2599 GTGCTGGAGCGAGGGCCCTTACACAGGTCTCTTCTGCGCCAGGAATACACGATCGTG 2658
QY 3496 AATGAGCTGCTGCTGACCAAGTGGAGGCTGGCTGCGCTGCGAGTGGAGGCTGCTC 3555
Db 2659 AATGAGCTGCTGCTGACCAAGTGGAGGCTGGCTGCGCTGCGAGTGGAGGCTGCTC 2718
QY 3556 CTGGGCGTGGCCACCGCGCCCGCAGAGGTCTCTTCCACGCTGCTCTCTCAACTTC 3615
Db 2719 CTGGGCGTGGCCACCGCGCCCGCAGAGGTCTCTTCCACGCTGCTCTCTCAACTTC 2778
QY 3616 ACCTACAGCCCCCGACACCAAGTCTCTGACATCTGCTGCTCTGCTGTTGATGGAGAGCAG 3675
Db 2779 ACCTACAGCCCCCGACACCAAGTCTCTGACATCTGCTGCTCTGCTGTTGATGGAGAGCAG 2838
QY 3676 TTCTCTGCTCAGCTGGGTGTA 3696
Db 2839 TTCTCTGCTCAGCTGGGTGTA 2859
```

RESULT 6

```
US-10-943-893-1
; Sequence 1, Application US/10943893
; Publication No. US20050058634A1
; GENERAL INFORMATION:
; APPLICANT: Zhiu, Yunxiang
; APPLICANT: Genzyme Corporation
; TITLE OF INVENTION: Methods for Introducing Mannose 6-Phosphate and Other Oligosaccharides
; TITLE OF INVENTION: onto Glycoproteins and its Application Thereof
; FILE REFERENCE: 07680.0031-01000
; CURRENT APPLICATION NUMBER: US/10/943,893
; PRIOR FILING DATE: 2004-09-20
; PRIOR APPLICATION NUMBER: US 10/051,711
; PRIOR FILING DATE: 2002-01-17
; PRIOR APPLICATION NUMBER: 60/263,078
; PRIOR FILING DATE: 2001-01-18
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 2877
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: recombinant human acid alpha glucosidase
US-10-943-893-1
```

```
Query Match 71.4%; Score 2643.4; DB 19; Length 2877;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 2650; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
```

1036 GAAACCGGTGTCACACCCCGGCCCTCCAGAGAGTGCACACAGTGCAGCGTCCCGCCC 1095
11 GATGCCCCAGGCACACCCCGGCCCTCCAGAGAGTGCACACAGTGCAGCGTCCCGCCC 269
1096 AACAGCGCGTTCGATTGGCGCCCTGACAAAGGCCATCACCCAGAGAAACAGTGCAGAGCCCGC 1155
270 AACAGCGCGTTCGATTGGCGCCCTGACAAAGGCCATCACCCAGAGAAACAGTGCAGAGCCCGC 329
1156 GCGTGGCTACATCCCTGCAAGCAGAGGGCTGCAGGGAGCCAGATGGGGCAGCCCTGG 1215
330 GCGTGGCTACATCCCTGCAAGCAGAGGGCTGCAGGGAGCCAGATGGGGCAGCCCTGG 389
1216 TGGCTTCTTCCACCCAGCTACCCAGCTACAGCTGAGAACTGAGCTCTCTGAAATG 1275
390 TGGCTTCTTCCACCCAGCTACCCAGCTACAGCTGAGAACTGAGCTCTCTGAAATG 449
1276 GGTACACCGGCCACCTGACCCGTAACAACCCACCTTCTTCCCAAGACATCTGAC 1335
450 GGTACACCGGCCACCTGACCCGTAACAACCCACCTTCTTCCCAAGACATCTGAC 509
1336 CTGCGGCTGAGCGTGATGAGAGCTGAGAACCGGCTCCATCTCACGATCAAGATCCA 1395
510 CTGCGGCTGAGCGTGATGAGAGCTGAGAACCGGCTCCATCTCACGATCAAGATCCA 569
1396 GCTAAACAGGCGCTACGAGGTGCCTTTGGAGACCCCGCGTGTCCACAGCCGGGCACCGTCC 1455
570 GCTAAACAGGCGCTACGAGGTGCCTTTGGAGACCCCGCGTGTCCACAGCCGGGCACCGTCC 629
1456 CCACTCTACAGCGTGAGTTCCTCGAGAGCCCTTCGGGGTATCTGTGACCCGAGCTG 1515
630 CCACTCTACAGCGTGAGTTCCTCGAGAGCCCTTCGGGGTATCTGTGACCCGAGCTG 689
1516 GACGGCGCGGTGCTGTGAAACAGACGAGTGGGGCCCTCTCTTTGCGGACAGTTCCTT 1575
690 GACGGCGCGGTGCTGTGAAACAGACGAGTGGGGCCCTCTCTTTGCGGACAGTTCCTT 749
1576 CAGCTGTCAACCTCGCTGCGCTCGCAGATATATCACAGGCGTGCAGGACCTCAGTCC 1635
750 CAGCTGTCCACCTCGCTGCGCTCGCAGATATATCACAGGCGTGCAGGACCTCAGTCC 809
1636 CTGATGCTCAGACACAGCTGAGACAGAGTACACCTGTGAAACCGGACCTTGGCGCCACG 1695
810 CTGATGCTCAGACACAGCTGAGACAGAGTACACCTGTGAAACCGGACCTTGGCGCCACG 869
1696 CCGGTGCGAACTCTACGGGTCTACCTTCTACCTCGCGCTGAGGACCGCGGGTGG 1755
870 CCGGTGCGAACTCTACGGGTCTACCTTCTACCTGGCGCTGAGGACCGCGGGTGG 929
1756 GCACAGGGGTGTCTCTGCTAAACAGCAATGCCATGGATGTGTCTGCAAGCGGAGCCCT 1815
930 GCACAGGGGTGTCTCTGCTAAACAGCAATGCCATGGATGTGTCTGCAAGCGGAGCCCT 989
1816 GCCCTTAGCTGGAGTTCGACAGGTGGATCTCTGATGTCTATCATCTTCTGGGCGCAGAG 1875
990 GCCCTTAGCTGGAGTTCGACAGGTGGATCTCTGATGTCTATCATCTTCTGGGCGCAGAG 1049
1876 CCAAGAGCGTGGTGCAGAGTACCTGAGCTGTGGGATACCCGTTCATGGCGCATAC 1935
1050 CCAAGAGCGTGGTGCAGAGTACCTGAGCTGTGGGATACCCGTTCATGGCGCATAC 1109
1936 TGGGGCTTGGGCTTCCACCTGTGCGCGCTGGGGTACTCTCTCAACCGCTATCACCCGCCAG 1995
1110 TGGGGCTTGGGCTTCCACCTGTGCGCGCTGGGGTACTCTCTCAACCGCTATCACCCGCCAG 1169
1996 GTGGTGGAGAACATGACAGGGCCCACTTCCCGCTTGGACGTCAATGGAGAACAGCTGGAC 2055
1170 GTGGTGGAGAACATGACAGGGCCCACTTCCCGCTTGGACGTCAATGGAGAACAGCTGGAC 1229
2056 TACATGAGCTCCGGAGGAGCTTACAGTTCACACAGGATGGCTTCCGGGAGCTTCCGGGCC 2115
1230 TACATGAGCTCCGGAGGAGCTTACAGTTCACACAGGATGGCTTCCGGGAGCTTCCGGGCC 1289

2116 ATGTCGACGAGAGCTGCACACAGGGCGCGCGCTCATATGATGATCGTGTGATCCTGCCATC 2175
1290 ATGGTGCAGGAGCTGCACACAGGGCGCGCGCTACATGATGATCGTGTGATCCTGCCATC 1349
2176 AGCAGCTCGGGCCCTGCGGGAGCTACAGGCCCTACGACGAGGCTCTGCGGAGGGGGTT 2235
1350 AGCAGCTCGGGCCCTGCGGGAGCTACAGGCCCTACGACGAGGCTCTGCGGAGGGGGTT 1409
2236 TTTCATCACCAACGAGACCGGCCAGCGCTGATTTGGGAAGGTATGCGCCGGTCCACTGCC 2295
1410 TTTCATCACCAACGAGACCGGCCAGCGCTGATTTGGGAAGGTATGCGCCGGTCCACTGCC 1469
2296 TTCCCCGACTTTCACCAACCCCAAGCCCTGCGCTGCGGAGGACATGCTGCTGAGTTC 2355
1470 TTCCCCGACTTTCACCAACCCCAAGCCCTGCGCTGCGGAGGACATGCTGCTGAGTTC 1529
2356 CATGACAGAGTGCCTTTCGACGGCTTGTGGATTGATGAAAGAGGCTTCCAACTTCATC 2415
1530 CATGACAGAGTGCCTTTCGACGGCTTGTGGATTGATGAAAGAGGCTTCCAACTTCATC 1589
2416 AGAGGCTCTGAGGACGGCTGCCCCAAATGAGCTGGAGAACCCACCTTACGTGCTGGG 2475
1590 AGGGCTCTGAGGACGGCTGCCCCAAATGAGCTGGAGAACCCACCTTACGTGCTGGG 1649
2476 GTGGTTGGGGGACCTTCCAGCGGCCCACTTCTGTGCTTCCAGCCACAGTTCCTCTCC 2535
1650 GTGGTTGGGGGACCTTCCAGCGGCCCACTTCTGTGCTTCCAGCCACAGTTCCTCTCC 1709
2536 ACACACTAACCTTGCAACCTCTACGGCTTGACCGAGAGGATCGCTTCCACAGGGGG 2595
1710 ACACACTAACCTTGCAACCTCTACGGCTTGACCGAGAGGATCGCTTCCACAGGGGG 1769
2596 CTGGTGAAGGCTCGGGGACACCGCCATTTGTGATCTCCCGCTCGACCTTTGCTGCGCAC 2655
1770 CTGGTGAAGGCTCGGGGACACCGCCATTTGTGATCTCCCGCTCGACCTTTGCTGCGCAC 1829
2656 GSCCGATACGCGGCCACTGGAACGGGACGTTGTGGAGCTCTCTGGAGACAGCTCGCTCC 2715
1830 GSCCGATACGCGGCCACTGGAACGGGACGTTGTGGAGCTCTCTGGAGACAGCTCGCTCC 1889
2716 TCCGTGCCAGAACTCTGAGTTTAACTTGTGGGGTGCCTCTGTTGGGGCCGACGCTC 2775
1890 TCCGTGCCAGAACTCTGAGTTTAACTTGTGGGGTGCCTCTGTTGGGGCCGACGCTC 1949
2776 TCGCGCTTCTGGGCAACACCTCAGAGGAGTGTGTGCGCTGAGACCCAGCTGGGGGCC 2835
1950 TCGCGCTTCTGGGCAACACCTCAGAGGAGTGTGTGCGCTGAGACCCAGCTGGGGGCC 2009
2836 TTCTACCCCTTCAATGCGGAAACCAACAGCCTGCTCAGTCTGCCCCAGAGCCGTACAGC 2895
2010 TTCTACCCCTTCAATGCGGAAACCAACAGCCTGCTCAGTCTGCCCCAGAGCCGTACAGC 2069
2896 TTCAGGAGCGGGCCAGAGGCTGAGAGGCGCTTCAACCTGGGCTACGCACTCCTC 2955
2070 TTCAGGAGCGGGCCAGAGGCTGAGAGGCGCTTCAACCTGGGCTACGCACTCCTC 2129
2956 CCCACCTCTACACTGTTCACAGCCCAACGCTGCGGGGAGACCGTGGCCCGGGCC 3015
2130 CCCACCTCTACACTGTTCACAGCCCAACGCTGCGGGGAGACCGTGGCCCGGGCC 2189
3016 CTCTTCTGGAGTTCCCAAGGACTTACGACCTGGAACCTGGAACCAACAGCTCTGTGG 3075
2190 CTCTTCTGGAGTTCCCAAGGACTTACGACCTGGAACCTGGAACCAACAGCTCTGTGG 2249
3076 GGGGAGCGGCTGCTCATCACCCAGTGTCCAGGCGGGGAGGCGGAGTGTGCTGCTAC 3135
2250 GGGGAGCGGCTGCTCATCACCCAGTGTCCAGGCGGGGAGGCGGAGTGTGCTGCTAC 2309
3136 TTCCCCCTTGGGACATGTTACGACCTGAGAGCGGTGCAATAGAGGCCCTTGGCAGCCTC 3195
2310 TTCCCCCTTGGGACATGTTACGACCTGAGAGCGGTGCAATAGAGGCCCTTGGCAGCCTC 2369
3196 CCACCCCACTGTCAGCTTCCCGGTGAGCCATCCACAGCAGGGGCGAGTGGGTGACG 3255


```

; PRIOR FILING DATE: 2003-06-20
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8
; LENGTH: 2937
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: RAP-1DU fusion sequence
US-10-812-849-8

Query Match      28.2%; Score 1045.8; DB 19; Length 2937;
Best Local Similarity 99.3%; Pred. No. 2.3e-253;
Matches 1050; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 CTTACCGCATGCGGGTTCGAGCGGGCTCTGTGGCTCTCTGCTCTGCGCACCGTG 60
DB |||
QY 4 CTTACCGCATGCGGGTTCGAGCGGGCTCTGTGGCTCTCTGCTCTGCGCACCGTG 63
DB |||
QY 61 CTCGGATCCTACTCGCGGAGAGAACACAGCCCAAGCCGTCGCCGAAACGCGAGTCCGGA 120
DB |||
QY 64 CTCGGATCCTACTCGCGGAGAGAACACAGCCCAAGCCGTCGCCGAAACGCGAGTCCGGA 123
DB |||
QY 121 GAGGAGTTCGCATGAGAGAGTTGAACCAAGCTGTGGGAGAGGCCCGAGCGATGCACTTT 180
DB |||
QY 124 GAGGAGTTCGCATGAGAGAGTTGAACCAAGCTGTGGGAGAGGCCCGAGCGATGCACTTT 183
DB |||
QY 181 CTTCCCGTGAGGCTGCGCGAGCTCCACGCTGTATCTGAAGATACAGGAGAGGACGAACCTC 240
DB |||
QY 184 CTTCCCGTGAGGCTGCGCGAGCTCCACGCTGTATCTGAAGATACAGGAGAGGACGAACCTC 243
DB |||
QY 241 GCGTGAAGAACTAAAGCTTGAACGCTTGGACGAGATGGGAGAGGAAGCGAGACTC 300
DB |||
QY 244 GCGTGAAGAACTAAAGCTTGAACGCTTGGACGAGATGGGAGAGGAAGCGAGACTC 303
DB |||
QY 301 ATACGCAACTCAATGCTATCTGGCCAAAGTATGCTCTGACGGAAGAAAGACGCTCGG 360
DB |||
QY 304 ATACGCAACTCAATGCTATCTGGCCAAAGTATGCTCTGACGGAAGAAAGACGCTCGG 363
DB |||
QY 361 CAGGTACAGCAACTCCCTCAGTGGCACCCAGGAAGCGGCTGATGACCCAGGCTG 420
DB |||
QY 364 CAGGTACAGCAACTCCCTCAGTGGCACCCAGGAAGCGGCTGATGACCCAGGCTG 423
DB |||
QY 421 GAAAGCTGTGGCACAAGCGAAGACCTCTCTGGGAAATTTCTCCGGCGAAGAACTGGACAAG 480
DB |||
QY 424 GAAAGCTGTGGCACAAGCGAAGACCTCTCTGGGAAATTTCTCCGGCGAAGAACTGGACAAG 483
DB |||
QY 481 CTTGCGGGAGTTCTCTGATCAAAAGAGAAAGTTTCAAGATACAACTCTCTGCTGGAG 540
DB |||
QY 484 CTTGCGGGAGTTCTCTGATCAAAAGAGAAAGTTTCAAGATACAACTCTCTGCTGGAG 543
DB |||
QY 541 ACCCTGAGCAGGACCGAAGAAATCCACGAGAGAGCTCATTTAGCCCTTCGGACCTTGAGCGAC 600
DB |||
QY 544 ACCCTGAGCAGGACCGAAGAAATCCACGAGAGAGCTCATTTAGCCCTTCGGACCTTGAGCGAC 603
DB |||
QY 601 ATCAAGGCGAGCGTCTCTGACAGAGGCAACGAGCTGAAGGAGAGCTGCGCAGCATC 660
DB |||
QY 604 ATCAAGGCGAGCGTCTCTGACAGAGGCAACGAGCTGAAGGAGAGCTGCGCAGCATC 663
DB |||
QY 661 AACCAGGCTTGAGACCGCTGCGCAGGCTGACGACACAGGCTACAGCACTGAGGCTGAG 720
DB |||
QY 664 AACCAGGCTTGAGACCGCTGCGCAGGCTGACGACACAGGCTTACAGCACTGAGGCTGAG 723
DB |||
QY 721 TTGAGGAGCCAGGGTGATTTGACCTGTGGGACCTGGCGAGTCCGCCAACTTCACGGAC 780
DB |||
QY 724 TTGAGGAGCCAGGGTGATTTGACCTGTGGGACCTGGCGAGTCCGCCAACTTCACGGAC 783
DB |||
QY 781 AAGGAGCTGGAGCGTTCGCGGAGGAGCTCAAGCACTTCAAGGCCAAAATTCGAGAGACAC 840
DB |||
QY 784 AAGGAGCTGGAGCGTTCGCGGAGGAGCTCAAGCACTTCAAGGCCAAAATTCGAGAGACAC 843
DB |||
QY 841 AACCCTACAGNAGCAGCTGGAGATTGCGCAGAGAACTGAGGCAACGAGGAGCGTG 900
DB |||

RESULT 9
US-10-600-862A-10
; Sequence 10, Application US/10600862A
; Publication No. US20050026823A1
; GENERAL INFORMATION:
; APPLICANT: Zankel et al.
; TITLE OF INVENTION: USE OF THE CHAPERONE RECEPTOR-ASSOCIATED PROTEIN (RAP) FOR THE
; DELIVERY OF THERAPEUTIC COMPOUNDS TO THE BRAIN AND OTHER TISSUES
; FILE REFERENCE: 30610/39383
; CURRENT APPLICATION NUMBER: US/10/600,862A
; NUMBER OF SEQ ID NOS: 28
; CURRENT FILING DATE: 2003-06-20
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 10
; LENGTH: 1398
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: RAP-GDNF fusion sequence
US-10-600-862A-10

Query Match      26.4%; Score 978.8; DB 19; Length 1398;
Best Local Similarity 99.3%; Pred. No. 1.6e-236;
Matches 983; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 59 TGCTCGGATCCTACTCGCGGAGAGAAACCCAGCCCAAGCCGTCGCCGAAACCGAGTCCG 118
DB |||
QY 2 TGGGGGGTTCTTACTCGCGGAGAGAAACCCAGCCCAAGCCGTCGCCGAAACCGAGTCCG 61
DB |||
QY 119 GAGAGAGTTCCGCAATGGAGAGTTGAACCAAGCTGTGGGAGAGGCCCGAGCACTGCATC 178
DB |||
QY 62 GAGAGAGTTCCGCAATGGAGAGTTGAACCAAGCTGTGGGAGAGGCCCGAGCACTGCATC 121
DB |||
QY 179 TTCCTCCCGTGAGGCTGCGCGAGCTCCAGCTGATCTGAAGATACAGGAGAGGAGCAAC 238
DB |||
QY 122 TTCCTCCCGTGAGGCTGCGCGAGCTCCAGCTGATCTGAAGATACAGGAGAGGAGCAAC 181
DB |||
QY 239 TCGCTCGAAGAAACTAAAGCTTTGACGCTTGAGAGAGATGGGAGAGAGGAGGAGAGAC 298
DB |||
QY 182 TCGCTCGAAGAAACTAAAGCTTTGACGCTTGAGAGAGATGGGAGAGAGGAGGAGAGAC 241
DB |||
QY 299 TCATACGCAACCTCAATGTCTCTTTGGCCAAAGTATGTTCTGGACGGAAGAGGAGCGTCT 358
DB |||
QY 242 TCATACGCAACCTCAATGTCTCTTTGGCCAAAGTATGTTCTGGACGGAAGAGGAGCGTCT 301
DB |||
QY 359 GGCAGGTACCAAGCAACTCCCTCAGTGGCAACCCAGAGAGAGCGGCTGGATGACCCAGGC 418
DB |||
QY 302 GGCAGGTACCAAGCAACTCCCTCAGTGGCAACCCAGAGAGAGCGGCTGGATGACCCAGGC 361
DB |||
QY 419 TCGAAAGCTGTGGCACAAGGCGAAGAGCTCTGGGAAATTTCTCCGCGAAGAACTGGACA 478
DB |||
QY 362 TCGAAAGCTGTGGCACAAGGCGAAGAGCTCTGGGAAATTTCTCCGCGAAGAACTGGACA 421
DB |||
QY 479 AGCTCTGGCGGAGTTCTTGCAATCAAAAGAGAAAGTTTCAAGATACAACTCTGCTGCG 538
DB |||
QY 422 AGCTCTGGCGGAGTTCTTGCAATCAAAAGAGAAAGTTTCAAGATACAACTCTGCTGCG 481
DB |||
QY 539 AGACCTGAGCAGGAGCCGAGAAATTCACAGGAAAGCTCATTTAGCCCTTCGAGCCTGAGCG 598
DB |||
```


482	Db		AGACCCCTGAGCAGGACCGGAAGAAATCCACGAGAACGTCATTAGCGCCCTCGGACCTGAGCG	541
599	Qy		ACATCAAGGGCAGCGTCTCTGCACACGAGGCACACGAGCTGAAGGAGAAAGCTGCGCAGCA	658
542	Db		ACATCAAGGGCAGCGTCTCTGCACACGAGGCACACGAGCTGAAGGAGAAAGCTGCGCAGCA	601
659	Qy		TCAACCAAGGGCTTGGACCGCTGCGCAGGGTCAGCCACCAAGGGCTACAGCACTGAGGCTG	718
602	Db		TCAACCAAGGGCTTGGACCGCTGCGCAGGGTCAGCCACCAAGGGCTACAGCACTGAGGCTG	661
719	Qy		AGTTTCGAGGAGCCACAGGGTGATTGACCTGTGGGACCTGGCGACCTGGCCAACTCTCAGG	778
662	Db		AGTTTCGAGGAGCCACAGGGTGATTGACCTGTGGGACCTGGCGACCTGGCCAACTCTCAGG	721
779	Qy		ACAAGGAGCTGGAGCGCTTCGCGGAGGAGCTCAAGACATTCGAAGCCAAATTCGAGAAGC	838
722	Db		ACAAGGAGCTGGAGCGCTTCGCGGAGGAGCTCAAGACATTCGAAGCCAAATTCGAGAAGC	781
839	Qy		ACAACCACTACACAAGCAGCTGAGGATTGCGCAGAGAACTGAGGCACGACGAGAGCG	898
782	Db		ACAACCACTACACAAGCAGCTGAGGATTGCGCAGAGAACTGAGGCACGACGAGAGCG	841
899	Qy		TGGCGACGCGACGGTGTAGCCGCGACGCCGAGAGACACGCCCTGCTGAGAGGGCGGA	958
842	Db		TGGCGACGCGACGGTGTAGCCGCGACGCCGAGAGACACGCCCTGCTGAGAGGGCGGA	901
959	Qy		CCAAGGAGCTGGGCTACACGGTGAAGAGCATCTGCGAGACCTGTCCGGCAGGATCTCCA	1018
902	Db		CCAAGGAGCTGGGCTACACGGTGAAGAGCATCTGCGAGACCTGTCCGGCAGGATCTCCA	961
1019	Qy		GAGTCCGCGCAGGCGAAGAAACCGGTGAC	1048
962	Db		GAGTCCGCGCGAGGCGAAGAAACCGGTTAC	991

```

RESULT 10
US-10-812-849-10
; Sequence 10, Application US/10812849
; Publication No. US20050042227A1
; GENERAL INFORMATION:
; APPLICANT: Zankel et al.
; TITLE OF INVENTION: MEGALIN-BASED DELIVERY OF THERAPEUTIC COMPOUNDS TO THE BRAIN AND
; TITLE OF INVENTION: OTHER TISSUES
; FILE REFERENCE: 30610/40037
; CURRENT APPLICATION NUMBER: US/10/812,849
; CURRENT FILING DATE: 2004-03-30
; PRIOR APPLICATION NUMBER: US 10/600,862
; PRIOR FILING DATE: 2003-06-20
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 10
; LENGTH: 1398
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: RAP-GDNF fusion sequence
US-10-812-849-10

```

Db	122	TTCTCCGCTGGAGGCTGGCCGAGCTCCA	CGCTGATCTGAAGATACAGGAGAGGAGCGAAC	181	
Qy	239	TCGCCCTGGAGAAGAACTAAAGCTTTGACGGCTTGGACGAAGATGGGAGAGAAAGAAAGCAGAC	298		
Db	182	TCGCCCTGGAGAAGAACTAAAGCTTTGACGGCTTGGACGAAGATGGGAGAGAAAGAAAGCAGAC	241		
Qy	299	TCATACGCAACCTCAATGTGTCCTTGGCCNAAGTATGGTCTGGACGGAAGAAAGAAAGAGCGCTC	358		
Db	242	TCATACGCAACCTCAATGTGTCCTTGGCCNAAGTATGGTCTGGACGGAAGAAAGAGAGCGCTC	301		
Qy	359	GGCAGGTGACCAAGCAACTCCCTCAGTGGCACCCAGGAAGA	CGGGCTGGATGACCCCAAGGC	418	
Db	302	GGCAGGTGACCAAGCAACTCCCTCAGTGGCACCCAGGAAGA	CGGGCTGGATGACCCCAAGGC	361	
Qy	419	TGGAAGAAGCTGTGGCA	CAGAGCGAAGACCTCTGGGAAA	TTCTCCGGGAAAGAACTGGACA	478
Db	362	TGGAAGAAGCTGTGGCA	CAGAGCGAAGACCTCTGGGAAA	TTCTCCGGGAAAGAACTGGACA	421
Qy	479	AGCTCTGGCGGGAGTTCTTGTCATCACAAAGAGAAA	AGTTACAGTAGTACACAGTCTCTGCTGG	538	
Db	422	AGCTCTGGCGGGAGTTCTTGTCATCACAAAGAGAAA	AGTTACAGTAGTACACAGTCTCTGCTGG	481	
Qy	539	AGACCTCTGACGAGGACCGAAGAAATCCA	CGAGAACTCAT	TAGCCCTCTGGACCTGAGCG	598
Db	482	AGACCTCTGACGAGGACCGAAGAAATCCA	CGAGAACTCAT	TAGCCCTCTGGACCTGAGCG	541
Qy	599	ACATCAAGGCGAGGCTCTGCACAGCAGGCACACGGAGCTGA	AGGAGAGAACTGGCAGCA	658	
Db	542	ACATCAAGGCGAGGCTCTGCACAGCAGGCACACGGAGCTGA	AGGAGAGAACTGGCAGCA	601	
Qy	659	TCAACAGGCGCTGGACCGCTCTCGCAGGGCTACGCCA	CCAGGGCTACAGCACTGAGGCTG	718	
Db	602	TCAACAGGCGCTGGACCGCTCTCGCAGGGCTACGCCA	CCAGGGCTACAGCACTGAGGCTG	661	
Qy	719	AGTTTCAGGAGGCCAGGGTGATTGA	CTGTGGGACCTTGGCGAGTCCGCCAACTCA	CGG	778
Db	662	AGTTTCAGGAGGCCAGGGTGATTGA	CTGTGGGACCTTGGCGAGTCCGCCAACTCA	CGG	721
Qy	779	ACAAGGAGCTGGAGGGCGTTCCGGGAGGAGCTCAAGCACTT	CGAAGCCAAATCGAGAAAGC	838	
Db	722	ACAAGGAGCTGGAGGGCGTTCCGGGAGGAGCTCAAGCACTT	CGAAGCCAAATCGAGAAAGC	781	
Qy	839	ACAACCACTACCAAGAGCAGCTGGAGATTGCGCACGAGAA	GCTGAGGCACGCAAGAGCG	898	
Db	782	ACAACCACTACCAAGAGCAGCTGGAGATTGCGCACGAGAA	GCTGAGGCACGCAAGAGCG	841	
Qy	899	TGGGCGACGGCGAGCGTGTGAGCCGACCGCGAGAAAGCA	CGCCCTGCTGGAGGGGCGGA	958	
Db	842	TGGGCGACGGCGAGCGTGTGAGCCGACCGCGAGAAAGCA	CGCCCTGCTGGAGGGGCGGA	901	
Qy	959	CCAAGGAGCTTGGGCTACACGGTGAAGAAGCATCTG	CAGGACCTGTCCGGCAGGATCTTCCA	1018	
Db	902	CCAAGGAGCTTGGGCTACACGGTGAAGAAGCATCTG	CAGGACCTGTCCGGCAGGATCTTCCA	961	
Qy	1019	GAGCTCGCGCCGAGGCGAGAAACCGGTGGAC	1048		
Db	962	GAGCTCGCGCCGAGGCGAGAAACCGGTGGAC	991		

RESULT 11
US-10-172-118-757
; Sequence 757, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients

Query Match	25.9%	Score 957	DB 17	Length 1493
Best Local Similarity	99.0%	Pred. No. 5.2e-231		
Matches 963	Conservative 0	Mismatches 10	Indels 0	Gaps 0
Qy	63	CGGATCCTACTCGCGGAGAGAACACAGCCGCTCCCGGAAACGCGAGTCGGGAGA	122	
Db	109	CGGCAAGTACTCGCGGAGAGAACACAGCCGCTCCCGGAAACGCGAGTCGGGAGA	168	
Qy	123	GGAGTTCCGATGAGAGAGTTGAACACAGCTGTGGAGAGGCCACGAGCTGCACTCTTCC	182	
Db	169	GGAGTTCCGATGAGAGAGTTGAACACAGCTGTGGAGAGGCCACGAGCTGCACTCTTCC	228	
Qy	183	TCCCGTCAGGCTGCGCGAGCTCCACGCTGATCTGAAGATACAGAGAGGGAGCAACTCGC	242	
Db	229	TCCCGTCAGGCTGCGCGAGCTCCACGCTGATCTGAAGATACAGAGAGGGAGCAACTCGC	288	
Qy	243	CTGGAAGAAACTAAAGCTTGA CGGCTTTGGACGAAGATGGGGAGAGGAAGCGGAGACTCAT	302	
Db	289	CTGGAAGAAACTAAAGCTTGA CGGCTTTGGACGAAGATGGGGAGAGGAAGCGGAGACTCAT	348	
Qy	303	ACGCAACTCAATGTCTATTGGGCCAAGTATGGTCTGGA CGGAAGAAAGAGAGCTCGGCA	362	
Db	349	ACGCAACTCAATGTCTATTGGGCCAAGTATGGTCTGGA CGGAAGAAAGAGAGCTCGGCA	408	
Qy	363	GGTGACAGCAACTCCCTCAGTGGCACCCACGAGAGAGGGGCTGGATGACCCACGAGCTGGA	422	
Db	409	GGTGACAGCAACTCCCTCAGTGGCACCCACGAGAGAGGGGCTGGATGACCCACGAGCTGGA	468	
Qy	423	AAAGCTGTGGCAAGGCGAAGACCTCTGGGAAATTTCTCCGGCGAAGAACTGGCAAGCT	482	
Db	469	AAAGCTGTGGCAAGGCGAAGACCTCTGGGAAATTTCTCCGGCGAAGAACTGGCAAGCT	528	
Qy	483	CTGCGGGAGTTCTTGCAATCAAAAGAGAAAGTTCA CGAGTACAACTGCTGTGGAGAC	542	
Db	529	CTGCGGGAGTTCTTGCAATCAAAAGAGAAAGTTCA CGAGTACAACTGCTGTGGAGAC	588	
Qy	543	CCTGAGCAGGACCGAAGAAATCCACGAGAAAGTCAATAGCCCTCGGACCTGAGCGGACAT	602	
Db	589	CCTGAGCAGGACCGAAGAAATCCACGAGAAAGTCAATAGCCCTCGGACCTGAGCGGACAT	648	
Qy	603	CAAGGGCAGCGTCTTGCAACAGCAGGCA CGAGCTGAAGAGAGAGCTGCGCAGCATCAA	662	
Db	649	CAAGGGCAGCGTCTTGCAACAGCAGGCA CGAGCTGAAGAGAGAGCTGCGCAGCATCAA	708	
Qy	663	CCAGGGCTGGA CCGCCTGCGCAGGGTCAAGCA CCGGGCTACAGCACTGAGGGCTGAGTT	722	
Db	709	CCAGGGCTGGA CCGCCTGCGCAGGGTCAAGCA CCGGGCTACAGCACTGAGGGCTGAGTT	768	
Qy	723	CGAGGAGCCAGGGTGATTGACTGTGGGA CTTGGCGAGCTCCGCAACCTTCAAGCAAA	782	
Db	769	CGAGGAGCCAGGGTGATTGACTGTGGGA CTTGGCGAGCTCCGCAACCTTCAAGCAAA	828	
Qy	783	GGAGCTGGAGCGTTCTGGGAGGAGCTCAAGCACTTCGAAGCAAAATCGAGAGACAA	842	
Db	829	GGAGCTGGAGCGTTCTGGGAGGAGCTCAAGCACTTCGAAGCAAAATCGAGAGACAA	888	
Qy	843	CCACTACAGAGCAGCTGGAGATTGGCACAGAGAGCTGAGGCAACGAGAGCGTGGG	902	

QY 423 AAAGCTGTGGCACAAGCGGAGAGCTCTGGGAATTTCTCGGCGAAGAACTGGACAAGCT 482
Db 469 AAAGCTGTGGCACAAGCGGAGAGCTCTGGGAATTTCTCGGCGAAGAACTGGACAAGCT 528
QY 483 CTGGCGGGAGTTCTCTGCATCACAAGAGAGAAAGTTTACAGAGTCAACAGTCTCTGTGGAGAC 542
Db 529 CTGGCGGGAGTTCTCTGCATCACAAGAGAGAAAGTTTACAGAGTCAACAGTCTCTGTGGAGAC 588
QY 543 CTGTAGCAGGACCGAAGAAATCCACGAGAAAGTCTATTAGCCCTCTCGGACCTGTAGCGACAT 602
Db 589 CTGTAGCAGGACCGAAGAAATCCACGAGAAAGTCTATTAGCCCTCTCGGACCTGTAGCGACAT 648
QY 603 CAAAGGACCGTCTCTGCACAGCAGGACACAGGAGCTGAAAGGAGAACTCGCAGAGTCAA 662
Db 649 CAAAGGACCGTCTCTGCACAGCAGGACACAGGAGCTGAAAGGAGAACTCGCAGAGTCAA 708
QY 663 CAAAGGACCGTCTCTGCACAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGG 722
Db 709 CAAAGGACCGTCTCTGCACAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGG 768
QY 723 CAAAGGACCGTCTCTGCACAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGG 782
Db 769 CAAAGGACCGTCTCTGCACAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGG 828
QY 783 CAAAGGACCGTCTCTGCACAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGG 842
Db 829 CAAAGGACCGTCTCTGCACAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGG 888
QY 843 CAAAGGACCGTCTCTGCACAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGG 902
Db 889 CAAAGGACCGTCTCTGCACAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGG 948
QY 903 CAAAGGACCGTCTCTGCACAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGG 962
Db 949 CAAAGGACCGTCTCTGCACAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGG 1008
QY 963 CAAAGGACCGTCTCTGCACAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTTCCAGAGC 1022
Db 1009 CAAAGGACCGTCTCTGCACAGGAGCTGAGGAGCTTCCAGAGCCTTCCAGAGCCTTCCAGAGC 1068
QY 1023 TCGCGCGGAGGCA 1035
Db 1069 TCGGCACAACGAA 1081

RESULT 13

US-10-342-887-757
; Sequence 757, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 757
; LENGTH: 1493
; TYPE: DNA
; ORGANISM: Homo sapiens

US-10-342-887-757

Query Match 25.9%; Score 957; DB 17; Length 1493;
Best Local Similarity 99.0%; Pred. No. 5.2e-231;
Matches 963; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 63 CGGATCCTACTCGCGGAGAGCAACAGCCCAAGCGTCCCGAAACGCGAGTCCGGAGA 122
Db 109 CGGCAAGTACTCGCGGAGAGAAACAGCCCAAGCGTCCCGAAACGCGAGTCCGGAGA 168
QY 123 GGAGTTCCCGCATGGAGAAAGTTGAAACAGCTGTGGGAGAAAGGCCAGCGACTGCTATTC 182
Db 169 GGAGTTCCCGCATGGAGAAAGTTGAAACAGCTGTGGGAGAAAGGCCAGCGACTGCTATTC 228
QY 183 TCCGCTGAGGCTGGCGCGAGCTCCAGCTGATCTGAAGATACAGGAGAGGAGCAATCTGC 242
Db 229 TCCGCTGAGGCTGGCGCGAGCTCCAGCTGATCTGAAGATACAGGAGAGGAGCAATCTGC 288
QY 243 CTGGAGAGAACTAAAGCTTGACGGCTTGACCAAGATGGGAGAGAGGAGCGAGACTCAT 302
Db 289 CTGGAGAGAACTAAAGCTTGACGGCTTGACCAAGATGGGAGAGAGGAGCGAGACTCAT 348
QY 303 ACGCAACCTCAATGTCATCTTGCCCAAGTATGCTCTGGACCGAAAGAAAGCAAGCTCGGCA 362
Db 349 ACGCAACCTCAATGTCATCTTGCCCAAGTATGCTCTGGACCGAAAGAAAGCAAGCTCGGCA 408
QY 363 GGTGACCAAGCAATCCCTCAGTGGCACCCAGAAAGACGGGCTGGATGACCCAGAGCTGGA 422
Db 409 GGTGACCAAGCAATCCCTCAGTGGCACCCAGAAAGACGGGCTGGATGACCCAGAGCTGGA 468
QY 423 AAAGCTGTGGCACAAGCGGAGAGCTCTGGGAAATTTCTCGGCGAAGAACTGGACAAGCT 482
Db 469 AAAGCTGTGGCACAAGCGGAGAGCTCTGGGAAATTTCTCGGCGAAGAACTGGACAAGCT 528
QY 483 CTGGCGGGAGTTCTCTGCATCACAAGAGAGAAAGTTTCAAGAGTACAACTCTCTGTGGAGAC 542
Db 529 CTGGCGGGAGTTCTCTGCATCACAAGAGAGAAAGTTTCAAGAGTACAACTCTCTGTGGAGAC 588
QY 543 CTGTAGCAGGACCGAAGAAATCCACGAGAACTCTATTAGCCCTCTGGACCTGTAGCGACAT 602
Db 589 CTGTAGCAGGACCGAAGAAATCCACGAGAACTCTATTAGCCCTCTGGACCTGTAGCGACAT 648
QY 603 CAAAGGACCGTCTCTGCACAGCAGGACACGCGAGCTGAAAGGAGAGCTCGGCGAGTCAA 662
Db 649 CAAAGGACCGTCTCTGCACAGCAGGACACGCGAGCTGAAAGGAGAGCTCGGCGAGTCAA 708
QY 663 CAAAGGACCGTCTCTGCACAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGG 722
Db 709 CAAAGGACCGTCTCTGCACAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGG 768
QY 723 CGAGGAGCCAGGGTGATTGACCTGTGGGACCTGTGGGACCTGTGGGACCTGTGGGACCTGTGG 782
Db 769 CGAGGAGCCAGGGTGATTGACCTGTGGGACCTGTGGGACCTGTGGGACCTGTGGGACCTGTGG 828
QY 783 CGAGGAGCCAGGGTGATTGACCTGTGGGACCTGTGGGACCTGTGGGACCTGTGGGACCTGTGG 842
Db 829 CGAGGAGCCAGGGTGATTGACCTGTGGGACCTGTGGGACCTGTGGGACCTGTGGGACCTGTGG 888
QY 843 CCAGTACCAAGAGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGG 902
Db 889 CCAGTACCAAGAGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGG 948
QY 903 CGAGGAGCCAGGGTGATTGACCTGTGGGACCTGTGGGACCTGTGGGACCTGTGGGACCTGTGG 962
Db 949 CGAGGAGCCAGGGTGATTGACCTGTGGGACCTGTGGGACCTGTGGGACCTGTGGGACCTGTGG 1008
QY 963 CGAGGAGCCAGGGTGATTGACCTGTGGGACCTGTGGGACCTGTGGGACCTGTGGGACCTGTGG 1022
Db 1009 CGAGGAGCCAGGGTGATTGACCTGTGGGACCTGTGGGACCTGTGGGACCTGTGGGACCTGTGG 1068
QY 1023 TCGCGCGGAGGCA 1035
Db 1069 TCGGCACAACGAA 1081

```
RESULT 14
US-10-472-516-3
; Sequence 3, Application US/10472516
; Publication No. US20040248785A1
; GENERAL INFORMATION:
; APPLICANT: Saenko, Evgheni L.
; APPLICANT: Sarafanov, Andrey G.
; TITLE OF INVENTION: Methods and Compositions for Reducing Heparan Sulfate
; TITLE OF INVENTION: Proteoglycan-Mediated Clearance of Factor VIII
; FILE REFERENCE: 1327.066001
; CURRENT APPLICATION NUMBER: US/10/472,516
; PRIOR FILING DATE: 2003-09-24
; PRIOR APPLICATION NUMBER: PCT/US02/00583
; PRIOR FILING DATE: 2002-01-11
; PRIOR APPLICATION NUMBER: US 60/260,904
; PRIOR FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 1493
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: (14)..(115)
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (116)..(1084)
; NAME/KEY: CDS
; LOCATION: (14)..(1084)
US-10-472-516-3

Query Match 25.9%; Score 957; DB 18; Length 1493;
Best Local Similarity 99.0%; Pred. No. 5.2e-231;
Matches 963; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 63 CGGATCTACTCGCGGAGAGAAACAGCCAGCCGCTCCCGAAACGGAGTCCGGAGA 122
DB 109 CGGCAAGTACTCGCGGAGAGAAACAGCCAGCCGCTCCCGAAACGGAGTCCGGAGA 168
QY 123 GGAGTTCCGATGGAGAAAGTTGAAACAGCTGTGGGAGAGGCGGACGACTGCATCTCC 182
DB 169 GGAGTTCCGATGGAGAAAGTTGAAACAGCTGTGGGAGAGGCGGACGACTGCATCTCC 228
QY 183 TCCGCTGAGGCTGGCGGAGCTCACGCTGATCTGAAGATACAGGAGAGGAGCAACTCGC 242
DB 229 TCCGCTGAGGCTGGCGGAGCTCACGCTGATCTGAAGATACAGGAGAGGAGCAACTCGC 288
QY 243 CTGGAGAGAACTAAAGCTTACGGCTTGGACGAAATGGGAGAGGAGGAGGAGCAACTCAT 302
DB 289 CTGGAGAGAACTAAAGCTTGGACGCTTGGAGAGATGGGAGAGGAGGAGGAGCAACTCAT 348
QY 303 ACGCAACCTCAATGTCTTGGCCCAAGTATGTCTGGACGGAAGAGGAGGAGGAGGAGGAG 362
DB 349 ACGCAACCTCAATGTCTTGGCCCAAGTATGTCTGGACGGAAGAGGAGGAGGAGGAGGAG 408
QY 363 GGTGACCAAGCACTCCCTCAGTGGGACCCAGGAGAGCGGGCTGGATGACCCAGGCTGGA 422
DB 409 GGTGACCAAGCACTCCCTCAGTGGGACCCAGGAGAGCGGGCTGGATGACCCAGGCTGGA 468
QY 423 AAAGCTGTGGCAACAGCGAAGCACTCTGGGAAATTTCTCCGGCGAAGCACTGGACAAGCT 482
DB 469 AAAGCTGTGGCAACAGCGAAGCACTCTGGGAAATTTCTCCGGCGAAGCACTGGACAAGCT 528
QY 483 CTGGCGGAGGTTCTCTGCATCAAAAAGAGAAAGTTTACAGAGTACAAAGTCTCTCTGGAGAC 542
DB 529 CTGGCGGAGGTTCTCTGCATCAAAAAGAGAAAGTTTACAGAGTACAAAGTCTCTCTGGAGAC 588
QY 543 CTTGAGCAGGAGCGGAAGAAATCCAGAGAAAGTCTTACCTAGCCCTCGGAGCTGAGGACAT 602
DB 543 CTTGAGCAGGAGCGGAAGAAATCCAGAGAAAGTCTTACCTAGCCCTCGGAGCTGAGGACAT 602

RESULT 15
US-10-741-600-440
; Sequence 440, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001499
; CURRENT APPLICATION NUMBER: US/10/741,600
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 73997
; SOFTWARE: PatSeq For Windows Version 4.0
; SEQ ID NO 440
; LENGTH: 2064
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-741-600-440

Query Match 25.8%; Score 954.2; DB 19; Length 2064;
Best Local Similarity 98.3%; Pred. No. 2.7e-230;
Matches 956; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

QY 63 CGGATCTACTCGCGGAGAGAAACAGCCAGCCGCTCCCGAAACGGAGTCCGGAGA 122
DB 146 CGGCAAGTACTCGCGGAGAGAAACAGCCAGCCGCTCCCGAAACGGAGTCCGGAGA 205
QY 123 GGAGTTCCGATGGAGAAAGTTGAAACAGCTGTGGGAGAGGCGGACGCACTGCATCTTCC 182
DB 206 GGAGTTCCGATGGAGAAAGTTGAAACAGCTGTGGGAGAGGCGGACGCACTGCATCTTCC 265
QY 183 TCCGCTGAGGCTGGCGGAGCTCCAGCTGATCTGAAGATACAGGAGAGGAGCAACTCGC 242
DB 266 TCCGCTGAGGCTGGCGGAGCTCCAGCTGATCTGAAGATACAGGAGAGGAGCAACTCGC 325
QY 243 CTGGAGAGAACTAAAGCTTACGGCTTGGACGAAATGGGAGAGGAGGAGGAGGAGGAGTCTAT 302
DB 326 CTGGAGAGAACTAAAGCTTACGGCTTGGACGAAATGGGAGAGGAGGAGGAGGAGGAGTCTAT 385
```

```
QY 303 ACCAAACCTCAATGTTCATCTTGCCCAAGTATGCTCTGGACGGAAAGAGGACGCTCGGCA 362
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
386 ACCARCCCTCARTGTTCATCTTGCCCAAGTATGCTCTGGACGGAAAGAGGACGCTCGGCA 445
QY 363 GGTGACCGCAACTCCCTCAGTGGCACCCAGGAGAGGGCTGGATGACCCAGGCTGGA 422
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
446 GGTGACCGCAACTCCCTCAGTGGCACCCAGGAGAGGGCTGGATGACCCAGGCTGGA 505
QY 423 AAAGCTGTGGCACAGCGCAAGACCTCTGGGAAATTCCTCGGCGAAAGAACTGGACAAAGCT 482
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
506 AAAGCTGTGGCACAGCGCAAGACCTCTGGGAAATTCCTCGGCGAAAGAACTGGACAAAGCT 565
QY 483 CTGGCGGGAGTTCTCTGCATCACAAGAGAAAGTTTCAAGTACAAAGTCTCTGCTGGAGAC 542
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
566 CTGGCGGGAGTTCTCTGCATCACAAGAGAAAGTTTCAAGTACAAAGTCTCTGCTGGAGAC 625
QY 543 CTTGACGCGGACCGAGAGAAATCCACAGAGACGTCATTAGCCCTCGGACCTGAGCGACAT 602
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
626 CTTGAGCGAGACCGAGAGAAATCCACAGAGACGTCATTAGCCCTCGGACCTGAGCGACAT 685
QY 603 CAAAGGCGAGCGTCTCTGCACAGCAGGACACGAGAGCTGAAGGAGAAAGCTGCGCAGCATCAA 662
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
686 CAAAGGCGAGCGTCTCTGCACAGCAGGACACGAGAGCTGAAGGAGAAAGCTGCGCAGCATCAA 745
QY 663 CCAGGGCCTGGACCCGCTCGCAGGGTCAAGCTACAGGGCTACAGCACTGAGGCTGAGTT 722
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
746 CCAGGGCCTGGACCCGCTCGCAGGGTCAAGCTACAGGGCTACAGCACTGAGGCTGAGTT 805
QY 723 CGAGGAGCCGAGGGTATGATTCCTGTGGACCTGGGCGAGTCCGCCAACCTCAGCGACAA 782
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
806 CGAGGAGCCGAGGGTATGATTCCTGTGGGACCTGGGCGAGTCCGCCAACCTCAGCGACAA 865
QY 783 GGAGCTGGAGGCGTTCGGGAGGAGCTCAAGCACTTCGAAGCCAAATCGAGAAAGCAAA 842
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
866 GGAGCTGGAGGCGTTCGGGAGGAGCTCAAGCACTTCGAAGCCAAATCGAGAAAGCAAA 925
QY 843 CCACTACCAAGACAGCTGGAGATTGCGCACGAGAAAGCTGAGGCAACGAGAGAGCGTGGG 902
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
926 CCACTACCAAGACAGCTGGAGATTGCGCACGAGAAAGCTGAGGCAACGAGAGAGYRTGG 985
QY 903 CGACGGCGAGCGTGTGAGCCGACGCGCGAGAAAGCAACGCCCTGCTGGAGGGCGGACCAA 962
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
986 CGAYGGCGAGCGTGTGAGCCGACGCGCGAGAAAGCAACGCCCTGCTGGAGGGCGGACCAA 1045
QY 963 GGAGCTGGGCTACACGCTGAAGAGCATCTGCAGGACCTGTCCGGCAGGATCTCCAGAGC 1022
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1046 GGAGCTGGGCTACACGCTGAAGAGCATCTGCAGGACCTGTCCGGCAGGATCTCCAGAGC 1105
QY 1023 TCGCGCCGAGGCA 1035
Db |||||:|||||
1106 TCGGCACACGAA 1118
```

Search completed: April 5, 2005, 00:59:58
Job time : 7246 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 4, 2005, 11:07:54 ; Search time 10485 Seconds
(without alignments)
13439.573 Million cell updates/sec

Title: US-10-600-862A-6
Perfect score: 3702
Sequence: 1 cttaccgcacgtcggggtcc.....tcagtggtgactcgag 3702

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: gb_est1:
2: gb_est2:
3: gb_est3:
4: gb_est4:
5: gb_est5:
6: gb_est6:
7: gb_gsa1:
8: gb_gsa2:
9: gb_gsa3:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2526.6	68.2	2859	9	AY413853 Homo sapi
2	1898.2	51.3	3488	3	AK088481 Mus muscu
3	1898.2	51.3	3493	3	AK052211 Mus muscu
4	1881.6	50.8	2859	9	AY413854 Pan trogl
5	1802.6	48.7	2862	9	AY413855 Mus muscu
6	959.8	25.9	1100	5	BM922403 AGENCOURT
7	957	25.9	1347	3	CR617514 full-leng
8	957	25.9	1373	3	CR603866 full-leng
9	957	25.9	1385	3	CR619388 full-leng
10	957	25.9	1392	3	CR606004 full-leng
11	957	25.9	1399	3	CR596948 full-leng
12	957	25.9	1399	3	CR625424 full-leng
13	957	25.9	1402	3	CR591537 full-leng
14	957	25.9	1409	3	CR619887 full-leng
15	957	25.9	1411	3	CR607197 full-leng
16	957	25.9	1418	3	CR616455 full-leng
17	957	25.9	1427	3	CR618610 full-leng
18	957	25.9	1430	3	CR607312 full-leng
19	957	25.9	1439	3	CR624552 full-leng
20	957	25.9	1442	3	CR625692 full-leng
21	957	25.9	1444	3	CR605976 full-leng
22	957	25.9	1449	3	CR597384 full-leng
23	957	25.9	1449	3	CR601788 full-leng
24	957	25.9	1449	3	CR611476 full-leng

25	957	25.9	1451	3	CR604735 full-leng
26	957	25.9	1451	3	CR625232 full-leng
27	957	25.9	1454	3	CR603821 full-leng
28	957	25.9	1456	3	CR591442 full-leng
29	957	25.9	1458	3	CR599568 full-leng
30	957	25.9	1460	3	CR604363 full-leng
31	957	25.9	1460	3	CR614498 full-leng
32	957	25.9	1463	3	CR626102 full-leng
33	957	25.9	1463	3	CR607263 full-leng
34	957	25.9	1465	3	CR613053 full-leng
35	927.2	25.0	1453	3	CR617840 full-leng
36	899.2	24.3	968	1	AL551402 full-leng
37	894.4	24.2	1004	5	BX337935 full-leng
38	884.6	23.9	1011	5	BX458886 full-leng
39	883	23.9	961	5	BX379126 full-leng
40	882	23.8	1003	1	AL518587 full-leng
41	880	23.8	1020	1	AL542049 full-leng
42	877.4	23.7	1013	1	AL518778 full-leng
43	877	23.7	1023	5	BX399454 full-leng
44	873.6	23.6	991	5	BX381548 full-leng
45	867.4	23.4	990	1	AL550423 full-leng

ALIGNMENTS

RESULT 1
AY413853
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
gene
ORIGIN

AY413853
Homo sapiens GAA gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
AY413853.1 GI:39769815
GSS.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2859)
Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejarawal, A.,
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
Adams, M.D. and Cargill, M.
Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
Science 302 (5652), 1960-1963 (2003)
14671302
Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejarawal, A.,
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
Adams, M.D. and Cargill, M.
Direct Submission
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
This sequence was made by sequencing genomic exons and ordering
them based on alignment.
Location/Qualifiers
1..2859
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..>2859
/gene="GAA"
/locus_tag="HCM5029"

Query Match 68.2%; Score 2526.6; DB 9; Length 2859;
Best Local Similarity 95.2%; Pred. No. 0;
Matches 2532; Conservative 0; Mismatches 127; Indels 0; Gaps 0;
QY 1036 GAAACGGTGCACACCCCGCGTCCCGAGAGAGTGCACAGTGCACGTCCCGCC 1095

[illegible]

QY 3256 CTGCGGCGCCCTGTGACACCATCAAGCTCCACCTCGGGCTGGGTACATCATCCCCCTG 3315
 Db 2419 CTGCGGCGCCCTGTGACACCATCAAGCTCCACCTCGGGCTGGGTACATCATCCCCCTG 2478
 QY 3316 CAGGGCCCTGGCTCACAACACAGAGTCCCGCAGCAGCCCATGCGCTGGCTGGCC 3375
 Db 2479 CAGGGCCCTGGCTCACAACACAGAGTCCCGCAGCAGCCCATGCGCTGGCTGGCC 2538
 QY 3376 CTAACCAAGGTTGAGAGGCGCCGAGGGAGCTTCTTGGACGATGGAGAGCCTGGAA 3435
 Db 2539 CTGACCAAGGTTGAGAGGCGCCGAGGGAGCTTCTTGGACGATGGAGAGCCTGGAA 2598
 QY 3436 GTGCTGGAGGAGGCGCTACACAGGTATCTTCTGCGCCAGGAATACAGATCGTG 3495
 Db 2599 GTGCTGGAGGAGGCGCTACACAGGTATCTTCTGCGCCAGGAATACAGATCGTG 2658
 QY 3496 AATGAGCTGTGCTGACCAAGTGAAGGAGCTGGCTCGAGCTGCAGAGGTGACTGTC 3555
 Db 2659 AATGAGCTGTGCTGACCAAGTGAAGGAGCTGGCTCGAGCTGCAGAGGTGACTGTC 2718
 QY 3556 CTGGGCGTGGCCACGCGCCCGACGACAGGTCTCTCCAAAGGTGTCTCTCCAACTTC 3615
 Db 2719 CTGGGCGTGGCCACGCGCCCGACGACAGGTCTCTCCAAAGGTGTCTCTCCAACTTC 2778
 QY 3616 ACTACAGCCCGACACCAAGTCTCGACATCTGTCTCGCTGTGATGGAGAGCAG 3675
 Db 2779 ACTACAGCCCGACACCAAGTCTCGACATCTGTCTCGCTGTGATGGAGAGCAG 2838
 QY 3676 TTCTCTGTCAGCTGGTGT 3694
 Db 2839 TTCTCTGTCAGCTGGTGT 2857

RESULT 2
 AK088481
 LOCUS
 DEFINITION Mus musculus 2 days neonate thymus thymic cells cDNA, RIKEN full-length enriched library, clone:E430018M07 product:glucosidase, alpha, acid, full insert sequence.

ACCESSION AK088481
 VERSION 1 GI:26353503
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1
 Carninci, P., and Hayashizaki, Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636

REFERENCE 2
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499374
 PUBMED 11042159

REFERENCE 3
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kusunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE 20530913
 PUBMED 11076861

REFERENCE
AUTHORSTITLE
JOURNALREFERENCE
AUTHORSTITLE
JOURNALREFERENCE
AUTHORSTITLE
JOURNALREFERENCE
AUTHORSTITLE
JOURNAL

COMMENT

FEATURES
source

CDS

4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)

5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)

6 (bases 1 to 3488)
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akashira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
 Direct Submission
 Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.jp/
 URL: http://fantom.gsc.riken.jp/
 Location/Qualifiers
 1. 3488
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="NOD"
 /db_xref="FANTOM DB:E430018M07"
 /db_xref="taxon:10090"
 /clone="E430018M07"
 /cell_type="thymic cells"
 /tissue_type="thymus"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stages="2 days neonate"
 237. 3098
 /note="unnamed protein product; glucosidase, alpha, acid (MGD|MG1:95609, GB|NM_008064, evidence: BLASTN, 99%, match=3299)
 putative"
 /codon_start=1
 /protein_id="BAC40382.1"
 /db_xref="GI:26353504"
 /translation="MNIRKPLCSNVGACTLISLTAVILGHLMLRELMLPQDLHESSGUNKTIRRHQGYKGPDLHQEQTEQPKAFTQCDVPPSPDFCDAPDGIGSEQCEARCGVPAQGVLEKQIQGFQFPSPYSPLENLSSTESGTLATLTSPTFFPCRVLTQLEVLMTDSRLHKIDPKASKRYEVPLETRVLSQAPSPLYSVSEEPFIVNRKLGKRVLLNTVAPLFFADQQLSTSLPSQHTGLGHLSPALMTDWARITLWNRDTPSQGTNLYGSHPFVLALEDGLAGHVFLLNSAMDVILQPSPALTWSTGGLDYYVFLGPEPKSVVOQYLDVVCYPPMPYMGJGFLCRWGYSTAIIVROVENMTKTHFPDQVMDLDTMDARRDTFNQDSFADFPDMVRHLHQGRYMMIVDPAISAGPAGSYRYPYDEGLRGVFTTNETGQLIGKVMFGTTAFPDFTTNPETLMDQMVSEPHA

(MGD|MGI:95609, GB|NM_008064, evidence: BLASTN, 99%,
match=3299)
putative"
/codon_start=1
/protein_id="BAC34888.1"
/db_xref="GI:26342452"
/translation="MNIRKPLCSNVVGACTILSLTAVILGHLMLRELMLLPQDLHE
SSGLWKTYPHQEGYKQPLHIOBTEQPKAEATQCDVPPSPSRFDCAKDGIQOEQ
CBARGCYVAGVKEPOIGOPWCFPPSPYSRLENLSSTESGYATLRTSTPTFF
PKDVLTLQLEVMETDSRLHFKIKOPASKRYEVPLETPRVLQAPSLYSVFSSEPF
GVVRRKLGRVLLNTVAPLFPADQFLQSLTSLPSOHTYGLGHELSPLMLSTDWARI
TLNDRTPPSQNLNLYSHPFYLALEDGLGHLGFLNSNAMDVILQSPALTWSTG
GILDVFLGPEKSVVQGLVDVGYPPFPYVGLFHLCRMGYSTALVRQVVENMT
RTHFLDVQNDLIDYMDARDFTFNQDSFADFPDMVRELHQDGRYVMIMVDAISSAG
PAGSYRYPDEGLRGGYFIETNGOPLIGKVPCTTAFPOFTNPETLDMQDMVSEFHA
OVPFDCMWMIDMEPSNVFSGSQGCPNNELENPYPVGVVGGILOAATICASSHOFLS
THYNLNLGLTEAIASSRALVKTGRTRFPFVSRSTPSGHGRYAGHWTDVDRSWEHL
AYSVDILQGLVGLVADICGFIIDTSEELCVRWITQLGAFYPMRNHNDLNSVPLQ
EYRFSETAQAMRKAFALRYALLPYLTFLPHRVGDTVARPLFLEPDEPSFRSA
DRQLMGAPALLITPVLEPKTEVTGYFPKGTWYNQMVSVDLSLTLPSPPSSFRSA
VOSKGOWLLEAPLDTINVHLREGYIIPQGPSLATTESRKOPMALAVALTAGSADG
ELFPWDGESLAVLERGAYTLVTFSAKNNTIVNKLVRVTEKGAELQLREVTVLGATAP
TQVLSNGIPVSNFTSPDNKSLAIPVSLMLGELFQISWS"
3469..3474
polyA_signal
/notes="putative"
3493
polyA_site
/notes="putative"

ORIGIN

Query Match 51.3%; Score 1898.2; DB 3; Length 3493;
Best Local Similarity 82.6%; Pred. No. 0;
Matches 2186; Conservative 0; Mismatches 458; Indels 3; Gaps 1;
QY 1048 CACCCGCGCGTCCAGAGCAGTCCACACAGTGGAGCTGCCCGCCCAACAGCGCTTC 1107
DB 452 CAGACTGAACAGCCCAAGAGCACCACACAGTGTGATGTGCCCCCAGCAGCCGCTTT 511
QY 1108 GATTGCGCCCTGACAGGCCATCACCCAGGAACAGTGGAGCGCCCGCGCTGCTGTAC 1167
DB 512 GACTGTGCCCGGACAAAGCATCTCACAGGAGCAATGCGAGCGCCGCGCTGTCTAT 571
QY 1168 ATCCCTGCAAGAGAGGGGTGAGGAGCGCCAGATGGGGAGCCCTGTGTCTTCCCA 1227
DB 572 GTCCAGCAGGCGAGTGTGTAAGGAGCGCAGATAGGGCAGCCCTGTGTCTTCCCT 631
QY 1228 CCCAGCTACCCAGCTACAGCTGAGAACCTGAGCTCTCTGAAATGGCTACACGGCC 1287
DB 632 CCAGCTACCCAGCTACCGTCTAGAGAACCTGAGCTCTACAGATCGGGGTACACAGCC 691
QY 1288 ACCCTGACCGGTACCAACCCCGACCTTTCCCAAGGACATCTGACCCCTGCGGCTGGAC 1347
DB 692 ACCCTGACCGGTACCAACCGCGACCTTTCCCAAGGATGTGTGACCTTACAGCTGGAG 751
QY 1348 GTGATGATGAGACTGAGAACCGCTTCCACTTACGATCAAGATCCAGCTAACAGGCGC 1407
DB 752 GTGCTGATGAGACAGACAGCGCCCTCCACTTCAAGATCAAGATCCTGTAGTAGCGC 811
QY 1408 TACGAGTGCCTTGAGAGCCCGCTGTCCACAGCGGGGACCGCTCCCACTCTACAGC 1467
DB 812 TACGAAGTGCCTTGAGAGCCCAACGTGTGCTGAGCAGGACCACTCCCACTTACAGC 871
QY 1468 GTGAGTGTTCGAGAGCGCTTGGGGTGATCGTGCAACGGCAGCTGAGCGCGCGCTG 1527
DB 872 GTGGAATTTTCAGAGGAACCCCTTGGAGTGATCGTTGGTAGGAAGCTTGTGTGGCGAGTG 931
QY 1528 CTGCTGAAACAGAGGTGGGGCCCTGTCTTTCGGACCACTGCTTTCAGCTGTCCAC 1587
DB 932 TTGCTGAACACAAACCGTGGCGCCCTGTCTTTCGCTGACCACTTCTGAGCTGTCCA 991
QY 1588 TCGCTGCCCTCGAGTATATACAGGCGCTCGCGAGCACTCAGTCCCCTGATGTCTCAGC 1647
DB 992 TCCCTGCCCTCCAGACATCAACAGGCTGGGGGAACACTCAGGCCCACTCATGTCTCAGC 1051

QY 1648 ACCAGCTGACACAGGATCACCCCTGTGTGAAACCGGAGCTTTCGCCCCACCGCTGCGAAC 1707
DB 1052 ACCGACTGGGCTCGTATCACCCCTCTGGAACCGGAGACACACCAACCCCTGCAAGGTACCAAC 1111
QY 1708 CTCTAGGGGTCTCACCCCTTTCTACCTGGCGCTGGAGGACGGGGTTCGGCACACAGGGGTG 1767
DB 1112 CTCTAGGGGTCTCACCTTTCTTACCTGGCACTGGAGGACGGGTGGCTTGGCTCACGGGTGTC 1171
QY 1768 TTCTCTGCTAAACAGCAATGSCCATGTGATGTGCTCTGACGCGAGCCCTGCCCTTACCTGG 1827
DB 1172 TTCTTGTCTAAACAGCAATGSCCATGTGATGTATCTCTGCAACCCAGCCAGCCCTAACCTGG 1231
QY 1828 AGGTCCACAGGTGGGATCCTTGGATCTTACATCTTCTTGGGGCCAGAGCCCAAGAGCTGG 1887
DB 1232 AGGTCAACGGGCGGATCCTTGGATGTGTATGTGTTCTTAGGCCAGAGCCCAAGAGCGTT 1291
QY 1888 GTGACGAGTACTTGGACGTTGTGGGATACCGTTTCATGCGCGCCATCTATGCGGCTTGGGC 1947
DB 1292 GTGCAACAATACCTGGATGTGTGGGATACCCCTTCTATGCGCTCCATCTATGCGGCTTGGGC 1351
QY 1948 TTCCACCTGTGCGCTGGGGCTTACTCTCCACCGCTATCACCCGCGCATGTTCGGCCAGGTAGTGAGAAC 2007
DB 1352 TTCCACCTGTGCGCTGGGGCTTACTCTCGACCGCCATGTTCGGCCAGGTAGTGAGAAC 1411
QY 2008 ATGACACAGGGCCCACTTCCCTCCCTGGAGCTCAATGGAACGACTGGAATACATGACTCC 2067
DB 1412 ATGACACAGGACACACTTCCCGCTGGAGCTGCAATGGAATGACTGGAATACATGACTCC 1471
QY 2068 CGGAGGGCTTCAAGCTTCAACAAAGATGGCTTCCGGGACTTCCCGGCCATGTCGAGGAG 2127
DB 1472 CGAAGAGACTTCACTTCAACAGCAGACAGCTTTCGGGACTTCCAGACATGTGTGGGGAG 1531
QY 2128 CTGCAACAGGGCGCGCGCTTACATGATGATCGTGGATCTCTGCCATCAGCAGCTCGGCG 2187
DB 1532 CTGCAACAGGATGGCGGGCTTACATGATGATCGTGGACCTTCCCATCAGCAGCGAGGC 1591
QY 2188 CTGCGGGGAGCTACAGGCCCTTACAGAGGCTCTCGGAGGGGGGTTTTCATACCAAC 2247
DB 1592 CTTGTGGGAGTTACAGGCCCTTACAGAGGCTCTCGGAGGGGGGTTTTCATACCAAC 1651
QY 2248 GAGACCGGCGAGCCCTGATTTGGGAAGGTATGGCCCGGGTCCACTGCTTCCCGCACTTC 2307
DB 1652 GAGACTGGGCGAGCCCTGATTTGGGAAGGTATGGCCCGGGTCCACTGCTTCCCGCACTTC 1711
QY 2308 ACCAACCACAGCCCTGGCTGTGGAGGACATGTGTGGCTGAGTTCCATGACCAGGTG 2367
DB 1712 ACCAACCCTGAGACCTTGGCTGTGGAGGACATGTGTGTGAGTTCCACCCAGGTG 1771
QY 2368 CCTTTCGACGGCTTGTGATTTGACATGAAAGAGCTTCCAACTTATCAGAGGCTCTGAG 2427
DB 1772 CCTTTCGATGGCATGTGGCTCGACATGAACGAAACCGTCCAACTTGTAGAGGCTCTCAG 1831
QY 2428 GACGGTCCCCCAACAATGAGCTGAGAACCCACCCCTAGCTGCGCTGGGGTGTGGGGGG 2487
DB 1832 CAGGGTCCCCCAACAATGAGCTGAGAACCCCCCTATGTGCCCGGGGTGTGGGGGG 1891
QY 2488 ACCCTCCAGGGCGCCCACTTGTGCTTCCAGCCACCACTTCTCTCCACACACTACAC 2547
DB 1892 ATCTTTCAGGGCAGCCCACTTGTGCTTCCAGCCCACTTCTCTCCACACACTACAC 1951
QY 2548 CTGCAACAACCTTACGGCTTACCGAGCATCGCTTCCACAGGGCGCTGTGAGGCT 2607
DB 1952 CTCCACAACCTTGTACGGCTTCACTGAAGCTATCGCTTCCAGCAGGGCCCTGTGTAAGCT 2011
QY 2608 CGGGGAGACAGCCCACTTGTGATCTCCCGCTCGACTTGTGCGGCGACCGCCGATACGCG 2667
DB 2012 CGGGGAGACAGCCCTTGTGATCTCCCGCTCAAACCTTCTCGGGGCAACCGCCGCTACGCT 2071
QY 2668 GGCCTCTGAGAGGGGAGCTGTGGAGCTCTTGGGAGAGCTCGCTTCTTCCCTGCTGCGAGAA 2727
DB 2072 GGTCACTGGAGGGGAGTGTGGAGCTCTTGGGAGGATCTTGTGATCTTGTGATCTCTGTGCCAGAC 2131
QY 2728 ATCTCTGCAAGTTTAACTGCTGGGGGTGCTCTGTGTCGGGGCGGACGTCTGCGGCTTCTCTG 2787

[illegible]

RESULT 4				
AY413854				
LOCUS	AY413854	2859 bp	DNA	linear
DEFINITION	Pan troglodytes GAA gene, VIRTUAL TRANSCRIPT, partial sequence, genomic survey sequence.			

ACCESSION	AY413854
VERSION	AY413854.1 GI:39769816
KEYWORDS	GSS.
SOURCE	Pan troglodytes (chimpanzee)
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE	1 (bases 1 to 2859)
AUTHORS	Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J., Adams,M.D. and Cargill,M.
TITLE	Infering nonneutral evolution from human-chimp-mouse orthologous gene trios
JOURNAL	Science 302 (5652), 1960-1963 (2003)
PUBMED	14671302
REFERENCE	2 (bases 1 to 2859)
AUTHORS	Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J., Adams,M.D. and Cargill,M.
TITLE	Direct Submission
JOURNAL	Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA
COMMENT	This sequence was made by sequencing genomic exons and ordering them based on alignment.
FEATURES	Location/Qualifiers
source	1..2859 /organism="Pan troglodytes" /mol_type="genomic DNA" /db_xref="taxon:9598" <1..>2859 /locus="GAA" /locus_tag="HCM5029"
gene	
ORIGIN	
Query Match	50.8%; Score 1881.6; DB 9; Length 2859;
Best Local Similarity	71.4%; Pred. No. 0;
Matches 1899; Conservative	0; Mismatches 760; Indels 0; Gaps 0;
QY	1036 GAATCCGGTGCACACCCCGCGCTCCAGAGCAGTGCCCAACAGTCGACGTCCGCCCCC 1095
DB	
QY	1096 AACAGCGCGTTTCATTGCGCCCTTGACAAGGCCCATCACCCAGGAACAGTGCAGGCGCCGC 1155
DB	
QY	1156 GGCTGCTGTACTATCCCTGCAAGACAGGGGCTGCAGGAGCCCATGAGTGGGGCAGCCCTGG 1215
DB	
QY	1216 TGCTTCTTCCACCAGCTACCCAGCTACAAGCTGGAGAACCTGAGTCTCTGAAATG 1275
DB	
QY	1276 GGCTACAGGCCAACCTGACCCGCTACCACCCCACCTTCTTCCCAAGGACATCTGACC 1335
DB	
QY	1336 CTGCGGCTGGACGTGATGATGGAGACTGTAGAACCGCGCTCCAATTACAGATCAAAGATCCA 1395
DB	
QY	1396 GCTAACAGGCGCTACAGGTGCCTTGGAGACCCCGCGTGTCCACAGCGGGCACCCTGCC 1455
DB	
QY	1456 CCATCTTACAGCTGAGTTCCTCCGAGGAGCCCTTCGGGGTGATCTGTCACCCGCGAGCTG 1515
DB	
QY	1516 GACGGCGCGGTGCTGCTGAACACGACGGTGGCGCCCTGTCTTTTGGGAGACGATGCTCTT 1575
DB	

RESULT 5	AY413855	2862 bp	DNA	linear	GSS 17-DEC-2003
LOCUS	Mus musculus GAA gene, VIRTUAL TRANSCRIPT, partial sequence,				
DEFINITION	genomic survey sequence.				
ACCESSION	AY413855				
VERSION	AY413855.1	GI:39769817			
KEYWORDS	GSS.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	1 (bases 1 to 2862) Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarimal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.				
TITLE	Infering nonneutral evolution from human-chimp-mouse orthologous gene trios				
JOURNAL	Science 302 (5652), 1960-1963 (2003)				
PUBMED	14671302				
REFERENCE	2 (bases 1 to 2862)				
AUTHORS	Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarimal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.				
TITLE	Direct Submission				
JOURNAL	Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA				
COMMENT	This sequence was made by sequencing genomic exons and ordering them based on alignment.				
FEATURES	Location/Qualifiers				
source	1..2862 /organism="Mus musculus" /mol_type="genomic DNA" /db_xref="taxon:10090" <1..>2862 /gene="GAA" /locus_tag="HOM5029"				
gene					
ORIGIN					
Query Match	48.7%;	Score 1802.6;	DB 9;	Length 2862;	
Best Local Similarity	78.7%;	Pred. No. 0;			
Matches 2082;	Conservative 0;	Mismatches 562;	Indels 3;	Gaps 1;	
Qy	1048	CACCCGGCGTCCAGACAGTGC	CGCCACACAGTGGACGTC	CCCCCCCCAACAGCGGCTTC	1107
Db	211	CAGACTGAACAGCCCAAGAAC	CCACACAGTGTGTGTGTC	CGCCCGCCAGACGCGCTTT	270
Qy	1108	GATTGGCCCTTGACAAGGCCAT	ACCAGGAACAGTCCGAGG	CGCCCGCGGCTGTGTCTAC	1167
Db	271	GACTGTGCCCCCGACAAAGCA	TCTACAGAGCAATCGAGG	CGCCCGCGGCTGTGTCTAT	330
Qy	1168	ATCCCTGCAAGACAGGGGCTG	CAGGGAGCCACAGATGGG	GCAGCCCTGTGTCTTCTCCA	1227
Db	331	GTCCACAGAGGCGAGTGTG	TGAGGAGCCACAGATAGG	GCAGCCCTGTGTCTTCTCC	390
Qy	1228	CCAGCTACCCAGCTACAAGCT	TGAGAACCTGTAGTCTCT	TGAAATGGGCTACACGGCC	1287
Db	391	CCAGCTACCCAGCTACCGCT	CTAGAACCTGTAGTCTCT	CACAGATCGGGGTACACAGCC	450
Qy	1288	ACCTGACCGGTACACCCGCT	CTCTCCACAGGATCCTGAC	CGCCCTGCGCTCGAC	1347
Db	451	ACCTGACCGGTACACCGCT	CTCTCTCCAAAGGATGTG	CTGACCTTACAGCTGGAG	510
Qy	1348	GTGATGAGGACTGAGAACCG	CTCTCACTTACAGATCAA	AGATCCAGCTAACAGGCGC	1407
Db	511	GTGCTGATGGAGACAGCGC	CTCTCACTTCAAGNTCAA	AGATCCCTGCTAGTAGGCG	570
Qy	1408	TACGAGGTGCCCTTGAGAC	CCCCCGGTGTGCCACAGC	CGGGCACCGTCCCCCTCTAC	1467
Qy	1468	GTGGAGTTCTCGAGAGAC	CCCTTCGGGGTGATCTGT	GACCGAGCTGGAGCGCCGGT	1521
Db	631	GTGGAGTTCTCAGAGGAA	CCCTTTGGAGTGATCTGT	TGTTAGGAAGCTTGGTGGC	690
Qy	1528	CTGCTGAACACAGCGTGG	CGCCCTCTTTCTTTTCG	GACACAGTTCCTTCACTGT	1587
Db	691	TTGCTGAACACAAACCGT	GGCCCCCTGTCTTCCG	CTGACAGTTCTTCGAGCTGT	750
Qy	1588	TGCTGCCCTCGCAGTATAT	CAACAGGCTCGCCGAGCA	CCCTCACTGTCCTGATCTC	1647
Db	751	TCCTGCCCTCCAGCATCA	CAGGCTCGGGGAAACAC	CTCAGCCCATCTGCTCAGC	810
Qy	1648	ACAGCTGGACAGGATCA	CCCTGTGTGAACCGGAC	CTTGGGCCACCGCCGGTGC	1707
Db	811	ACCGACTGGGCTCGTAT	CACCTCTGTGAACCGG	GACACACACCTCTGCAAG	870
Qy	1708	CTCTACGGGTCTCACCT	TTTCTACTCGCGCTGAG	GACCGCGGTTCGGCACAC	1767
Db	871	CTCTACGGGTCACTCT	TTTCTACTCGCGCTGAG	GACCGGTTCGGCTCAGCGT	930
Qy	1768	TTCTCTGCTAAACAGCA	ATGCTGATGTCTCTG	CAGCGCCCTGCTTTCGCTG	1827
Db	931	TTCTGCTTAACAGCAAT	GATGATCTCTGCAAC	CCAGCCCTTAACCTGCG	990
Qy	1828	AGGTCAACAGGTGGAT	CTCTGATCTTCTATCT	TTCGCGCCACAGAGCGGT	1887
Db	991	AGGTCAACAGGTGGAT	CTCTGATCTTCTATCT	TTCGCGCCACAGAGCGGT	1050
Qy	1888	GTGACAGTACCTGGAG	TTTGGGATACCCGTT	CTATGCGGCCATCTTGGG	1947
Db	1051	GTGCAACAATCTGGAT	TTTGGGATACCCCT	TTCATGCTTCCATCTCGG	1110
Qy	1948	TTCCACCTGTGCGCTGG	GGCTACTCTCTCACCG	CTATCACCGCCAGGTGTG	2007
Db	1111	TTCCACCTGTGCGCTGG	GGCTACTCTCTCACCG	CTATCACCGCCAGGTGTG	1170
Qy	2008	ATGACAGGGGCCACTT	CCCCCTCGGACGTCCA	ATGGAACGCTTGGACTAT	2067
Db	1171	ATGACAGGGACACACT	TTCCCGCTGGACGTG	CAATGGAAATGACCTGG	1230
Qy	2068	CGAGGAGCTTCAAGT	TGCTTCCGGGACTT	CCCGGCCATGTTGTCAGG	2127
Db	1231	CGAAGAGACTTCACT	TTTCAACACAGACAG	CTTTCCCGACTTCCGAG	1290
Qy	2128	CTGACACAGGGCGCG	CGCTACTATGATCTGT	GATCTCTGCATCACAG	2187
Db	1291	CTGACACAGGATGG	CGCGCTCTATGATCT	GTGACCTGCCATCAG	1350
Qy	2188	CCTGCGGGAGCTACAG	CGCCCTTACAGAGGG	GTCTGCGAGGGGGTTTT	2247
Db	1351	CCTGCTGGAGTTACAG	GGCCCTACAGAGGG	GTCTGCGAGGGGGTGTT	1410
Qy					

Db	697	CCAGGGCTTGAGACCGCCTGCGCAGGGTCAAGCCACAGGGCTACAGCACTGAGGCTGAGTT	756
QY	723	CGAGGAGCCCGAGGGTGATTGACCTGTGGGACCTGGCGCAGTCCGCCAACCTCACGGACAA	782
Db	757	CGAGGAGCCCGAGGGTGATTGACCTGTGGGACCTGGCGCAGTCCGCCAACCTCACGGACAA	816
QY	783	GGAGCTGGAGGGCTTCGGGAGGAGCTCAAGCACTTCGAAGCCAAATCGAGAAGCACAA	842
Db	817	GGAGCTGGAGGGCTTCGGGAGGAGCTCAAGCACTTCGAAGCCAAATCGAGAAGCACAA	876
QY	843	CCACTACCAAGACAGCTGGAGTTCGGCACGAGAGCTGAGGCACGCGAGAGCGCTGGG	902
Db	877	CCACTACCAAGACAGCTGGAGTTCGGCACGAGAGCTGAGGCACGCGAGAGCGCTGGG	936
QY	903	CGACGCGCAGCGTGTGAGCGCGCAGCGCAGAGACGCGCCCTGCTGGAGGGCGGACCAA	962
Db	937	CGACGCGCAGCGTGTGAGCGCGCAGCGCAGAGACGCGCCCTGCTGGAGGGCGGACCAA	996
QY	963	GGAGCTGGGCTACACGGTGAAGAAGCATCTGCAGGACCTGTCCGGCAGGATCTCCAGAGC	1022
Db	997	GGAGCTGGGCTACACGGTGAAGAAGCATCTGCAGGACCTGTCCGGCAGGATCTCCAGAGC	1056
QY	1023	TCGCGCCGAGGCA	1035
Db	1057	TCGGCACAAACGAA	1069
RESULT 8			
LOCUS			
DEFINITION			
full-length cDNA clone CS0D1004YE06 of Placenta Cot 25-normalized			
of Homo sapiens (human).			
ACCESSION			
CR603866			
VERSION			
CR603866.1 GI:50484673			
KEYWORDS			
HTC; CNS1T_cDNA			
SOURCE			
Homo sapiens			
ORGANISM			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE			
1 (bases 1 to 1373)			
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.			
Full-length cDNA libraries and normalization			
Unpublished			
TITLE			
Full-length cDNA libraries and normalization			
REMARK			
Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/Invitrogen Corporation 1600 Faraday Avenue			
REFERENCE			
2 (bases 1 to 1373)			
Genoscope.			
AUTHORS			
Direct Submission			
TITLE			
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :			
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr			
- Web : www.genoscope.cns.fr)			
COMMENT			
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.			
FEATURES			
Location/Qualifiers			
source			
1..1373			
/organism="Homo sapiens"			
/mol_type="mRNA"			
/db_xref="taxon:9606"			
/clone="CS0D1004YE06"			
/tissue_type="Placenta Cot 25-normalized"			
/plasmid="pCMVSPORT_6"			
ORIGIN			
Query Match			
Best Local Similarity 99.0%; Pred. No. 1.2e-185;			
Matches 963; Conservative 0; Mismatches 10; Indels 0; Gaps 0;			
QY	63	CGGATCCTACTCGCGGAGAGAACCGCCCAAGCCGTCGCCGAACGGAGTCCGGAGA	122


```

VERSION      CR619388.1  GI:50500195
KEYWORDS     HTC; CNSLT_CDNA.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
             Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 1385)
AUTHORS      Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE        Full-length cDNA libraries and normalization
JOURNAL      Unpublished
REMARK       Contact : Feng Liang Email : fliang@lifetech.com URL :
             http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
             Faraday Avenue
REFERENCE    2 (bases 1 to 1385)
AUTHORS      Genoscope.
TITLE        Direct Submission
JOURNAL      Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
             BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
             - Web : www.genoscope.cns.fr)
COMMENT      1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
             end enriched, double-strand cDNA was digested with Not I and cloned
             into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
             was normalized. Library was constructed by Life Technologies, a
             division of Invitrogen.
FEATURES     Location/Qualifiers
             source
               1..1385
               /organism="Homo sapiens"
               /mol_type="mRNA"
               /db_xref="taxon:9606"
               /clone="CS0D1011YN09"
               /tissue_type="Placenta Cot 25-normalized"
               /plasmid="pCMVSPORT_6"
ORIGIN
Query Match      25.98; Score 957; DB 3; Length 1385;
Best Local Similarity 99.08; Pred. No. 1.2e-185;
Matches 963; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY      63  CGGATCTCTACTCGGGGAGAGAACACAGCCCAAGCCGCTCCCGAAACCGGAGTCCGGAGA 122
DB      630  CAAGGGCAGCGTCTCTGCACAGCAGGACACACGGAGCTGAAGGAAAGCTGCGCAGCATCAA 662
QY      90  CGGCAGTACTCGGGGAGAGAACACAGCCCAAGCCGCTCCCGAAACCGGAGTCCGGAGA 149
DB      630  CAAGGGCAGCGTCTCTGCACAGCAGGACACACGGAGCTGAAGGAAAGCTGCGCAGCATCAA 669
QY      123  GGAATTCGCGATGAGAGAGTTGAAACCAAGCTGTGGGAGAGGCCAGGCACTGCATCTTCC 182
DB      630  CAAGGGCAGCGTCTCTGCACAGCAGGACACACGGAGCTGAAGGAAAGCTGCGCAGCATCAA 672
QY      150  GGAATTCGCGATGAGAGAGTTGAAACCAAGCTGTGGGAGAGGCCAGGCACTGCATCTTCC 209
DB      630  CAAGGGCAGCGTCTCTGCACAGCAGGACACACGGAGCTGAAGGAAAGCTGCGCAGCATCAA 675
QY      183  TCCGCTGAGCTGGCGAGCTCCACCTGATCTGAAGATACAGGAGGAGGAGCAACTCGC 242
DB      630  CAAGGGCAGCGTCTCTGCACAGCAGGACACACGGAGCTGAAGGAAAGCTGCGCAGCATCAA 678
QY      210  TCCGCTGAGCTGGCGAGCTCCACCTGATCTGAAGATACAGGAGGAGGAGCAACTCGC 269
DB      630  CAAGGGCAGCGTCTCTGCACAGCAGGACACACGGAGCTGAAGGAAAGCTGCGCAGCATCAA 681
QY      243  CTGGAAGAACTAAAGCTTGACGGCTTGACGAAGATGGGGAGAGGAGGAGGAGCACTCAT 302
DB      630  CAAGGGCAGCGTCTCTGCACAGCAGGACACACGGAGCTGAAGGAAAGCTGCGCAGCATCAA 684
QY      270  CTGGAAGAACTAAAGCTTGACGGCTTGACGAAGATGGGGAGAGGAGGAGGAGCACTCAT 329
DB      630  CAAGGGCAGCGTCTCTGCACAGCAGGACACACGGAGCTGAAGGAAAGCTGCGCAGCATCAA 687
QY      303  ACSCAACCTCAATGTATCTTGGCCCAAGTATGTCTGGACGGAAAGAGGAGCCTCGGCA 362
DB      630  CAAGGGCAGCGTCTCTGCACAGCAGGACACACGGAGCTGAAGGAAAGCTGCGCAGCATCAA 690
QY      330  ACSCAACCTCAATGTATCTTGGCCCAAGTATGTCTGGACGGAAAGAGGAGCCTCGGCA 389
DB      630  CAAGGGCAGCGTCTCTGCACAGCAGGACACACGGAGCTGAAGGAAAGCTGCGCAGCATCAA 692
QY      363  GGTGACGAGCACTCCCTAGTGGCCACCCAGGAGAGCGGCTGGATGACCCCGGCTGGA 422
DB      630  CAAGGGCAGCGTCTCTGCACAGCAGGACACACGGAGCTGAAGGAAAGCTGCGCAGCATCAA 695
QY      390  GGTGACGAGCACTCCCTAGTGGCCACCCAGGAGAGCGGCTGGATGACCCCGGCTGGA 449
DB      630  CAAGGGCAGCGTCTCTGCACAGCAGGACACACGGAGCTGAAGGAAAGCTGCGCAGCATCAA 698
QY      423  AAAGCTGTGGCACAAGCGGAAGACCTCTGGGAAATTTCTCGGGGAGAGAACTGGACAAAGCT 482
DB      630  CAAGGGCAGCGTCTCTGCACAGCAGGACACACGGAGCTGAAGGAAAGCTGCGCAGCATCAA 701
QY      450  AAAGCTGTGGCACAAGCGGAAGACCTCTGGGAAATTTCTCGGGGAGAGAACTGGACAAAGCT 509
DB      630  CAAGGGCAGCGTCTCTGCACAGCAGGACACACGGAGCTGAAGGAAAGCTGCGCAGCATCAA 704
QY      483  CTGGCGGGAGTTCTGTCATCAAAAGAGAAAGTTTCAAGATACAAAGCTCTCTGCTGGAGAC 542
DB      630  CAAGGGCAGCGTCTCTGCACAGCAGGACACACGGAGCTGAAGGAAAGCTGCGCAGCATCAA 707
QY      510  CTGGCGGGAGTTCTGTCATCAAAAGAGAAAGTTTCAAGATACAAAGCTCTCTGCTGGAGAC 569
DB      630  CAAGGGCAGCGTCTCTGCACAGCAGGACACACGGAGCTGAAGGAAAGCTGCGCAGCATCAA 710
QY      543  CTTGAGCAGGACCGAGAGAAATCCACGAGAAAGCTATAGCCCTCTCGGACCTGAGCGACAT 602
DB      630  CAAGGGCAGCGTCTCTGCACAGCAGGACACACGGAGCTGAAGGAAAGCTGCGCAGCATCAA 713
QY      570  CTTGAGCAGGACCGAGAGAAATCCACGAGAAAGCTATAGCCCTCTCGGACCTGAGCGACAT 629
DB      630  CAAGGGCAGCGTCTCTGCACAGCAGGACACACGGAGCTGAAGGAAAGCTGCGCAGCATCAA 716
QY      603  CAAGGGCAGCGTCTCTGCACAGCAGGACACACGGAGCTGAAGGAAAGCTGCGCAGCATCAA 662
DB      630  CAAGGGCAGCGTCTCTGCACAGCAGGACACACGGAGCTGAAGGAAAGCTGCGCAGCATCAA 669
QY      663  CAAGGGCAGCGTCTCTGCACAGCAGGACACACGGAGCTGAAGGAAAGCTGCGCAGCATCAA 722
DB      630  CAAGGGCAGCGTCTCTGCACAGCAGGACACACGGAGCTGAAGGAAAGCTGCGCAGCATCAA 729
QY      723  CGAGGAGCCAGGGTGATTGACCTGTGGGACCTTGGGCGAGTTCGCCCAACCTCTACGAGCAAA 782
DB      630  CAAGGGCAGCGTCTCTGCACAGCAGGACACACGGAGCTGAAGGAAAGCTGCGCAGCATCAA 735
QY      750  CGAGGAGCCAGGGTGATTGACCTGTGGGACCTTGGGCGAGTTCGCCCAACCTCTACGAGCAAA 809
DB      630  CAAGGGCAGCGTCTCTGCACAGCAGGACACACGGAGCTGAAGGAAAGCTGCGCAGCATCAA 812
QY      783  GGAGCTGGAGGGCTTCCGGGAGGAGCTCAAGCACTTTCGAAGCCAAATCGAGAGACACAA 842
DB      630  CAAGGGCAGCGTCTCTGCACAGCAGGACACACGGAGCTGAAGGAAAGCTGCGCAGCATCAA 845
QY      810  GGAGCTGGAGGGCTTCCGGGAGGAGCTCAAGCACTTTCGAAGCCAAATCGAGAGACACAA 869
DB      630  CAAGGGCAGCGTCTCTGCACAGCAGGACACACGGAGCTGAAGGAAAGCTGCGCAGCATCAA 872
QY      843  CCNCTACCAAGAGCAGCTGGAGATTTCGCCAGAGAGCTGAGCAGCAGCAGAGCGTGGG 902
DB      630  CAAGGGCAGCGTCTCTGCACAGCAGGACACACGGAGCTGAAGGAAAGCTGCGCAGCATCAA 905
QY      870  CCNCTACCAAGAGCAGCTGGAGATTTCGCCAGAGAGCTGAGCAGCAGCAGAGCGTGGG 929
DB      630  CAAGGGCAGCGTCTCTGCACAGCAGGACACACGGAGCTGAAGGAAAGCTGCGCAGCATCAA 932
QY      903  CGAGCGGAGCGTGTGAGCGCGCAGCGCAGAGAGACGCGCTCTGGAGGGCGGAGCAAA 962
DB      630  CAAGGGCAGCGTCTCTGCACAGCAGGACACACGGAGCTGAAGGAAAGCTGCGCAGCATCAA 965
QY      930  CGAGCGGAGCGTGTGAGCGCGCAGCGCAGAGAGACGCGCTCTGGAGGGCGGAGCAAA 989
DB      630  CAAGGGCAGCGTCTCTGCACAGCAGGACACACGGAGCTGAAGGAAAGCTGCGCAGCATCAA 992
QY      963  GGAGCTGGGCTACACGGTGAAGAGCATCTGCAGGACCTGTTCGGGAGGATCTCCAGAGC 1022
DB      630  CAAGGGCAGCGTCTCTGCACAGCAGGACACACGGAGCTGAAGGAAAGCTGCGCAGCATCAA 1025
QY      990  GGAGCTGGGCTACACGGTGAAGAGCATCTGCAGGAGCCTGTTCGGGAGGATCTCCAGAGC 1049
DB      630  CAAGGGCAGCGTCTCTGCACAGCAGGACACACGGAGCTGAAGGAAAGCTGCGCAGCATCAA 1052
QY      1023  TCGCGCCGAGGCA 1035
DB      630  CAAGGGCAGCGTCTCTGCACAGCAGGACACACGGAGCTGAAGGAAAGCTGCGCAGCATCAA 1038
QY      1050  TCGGCACAACGAA 1062
DB      630  CAAGGGCAGCGTCTCTGCACAGCAGGACACACGGAGCTGAAGGAAAGCTGCGCAGCATCAA 1065

```

```
Query Match      25.9%; Score 957; DB 3; Length 1392;
Best Local Similarity 99.0%; Pred. No. 1.2e-185;
Matches 963; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 63 CGGATCTACTCGGGGAGAGAACACGACCCCAAGCGTCCCGCAAAACGGAGTCCGGAGA 122
DB 41 CGGAAAGTACTCGCGGGAGAGAACACGACCCCAAGCGTCCCGCAAAACGGAGTCCGGAGA 100
QY 123 GGAGTTCCGATCGAGAGTTGAACACAGCTGTGGGAGAGAGCCACGACGATGCTCTTC 182
DB 101 GGAGTTCCGATCGAGAGTTGAACACAGCTGTGGGAGAGAGCCACGACGATGCTCTTC 160
QY 183 TCCCGTAGGCTGCCGAGCTCCACGCTGATCTGAAGATACAGGAGAGGAGCGAACTCGC 242
DB 161 TCCCGTAGGCTGCCGAGCTCCACGCTGATCTGAAGATACAGGAGAGGAGCGAACTCGC 220
QY 243 CTGGAAGAACTAAAGCTTGACCGCTTGACGAGATGGGAGAGAGGAGAGGAGGAGTCTAT 302
DB 221 CTGGAAGAACTAAAGCTTGACCGCTTGACGAGATGGGAGAGAGGAGAGGAGGAGTCTAT 280
QY 303 ACGCAACCTCAATGTCTATCTTGGCCCAAGTATGTCTGGACGGAAGAGAGCGCTCGCA 362
DB 281 ACGCAACCTCAATGTCTATCTTGGCCCAAGTATGTCTGGACGGAAGAGAGCGCTCGCA 340
QY 363 GGTGACCAAGCACTCCCTCAGTGGCAACCCAGGAAGACGGGCTGGATGACCCCAAGGCTGA 422
DB 341 GGTGACCAAGCACTCCCTCAGTGGCAACCCAGGAAGACGGGCTGGATGACCCCAAGGCTGA 400
QY 423 AAGCTGTGGCAAGCGGAGAGCTCTGGGAATTTCTCGGCGAGAGAACTGGAGCAAGCT 482
DB 401 AAGCTGTGGCAAGCGGAGAGCTCTGGGAATTTCTCGGCGAGAGAACTGGAGCAAGCT 460
QY 483 CTGCGCGGAGTCTCTGCATCACAAAGAGAAAGTTTACAGAGTCAACAGTCTCTGCTGGAGAC 542
DB 461 CTGCGCGGAGTCTCTGCATCACAAAGAGAAAGTTTACAGAGTCAACAGTCTCTGCTGGAGAC 520
QY 543 CTTGAGCAGGACCGAAGAAATCCACAGAGAACCGTATTAGCCCCCTCGGACCTTGAGCGACAT 602
DB 521 CTTGAGCAGGACCGAAGAAATCCACAGAGAACCGTATTAGCCCCCTCGGACCTTGAGCGACAT 580
QY 603 CAGGCGCGGCTCTCTGCACAGCAGGACACACGGAGCTGAAGAGAGAGCTGGGCGAGATCAA 662
DB 581 CAGGCGCGGCTCTCTGCACAGCAGGACACACGGAGCTGAAGAGAGAGCTGGGCGAGATCAA 640
QY 663 CAGGCGCTGAGCGCTGCGCAGGCTCAGCCACAGGCTACAGCACTGAGGCTGAGTT 722
DB 641 CAGGCGCTGAGCGCTGCGCAGGCTCAGCCACAGGCTACAGCACTGAGGCTGAGTT 700
QY 723 CGAGGAGCCCGAGGTTGATTGA CTTGTGGGACTTGGCGCAGTCCGCCA CTTCA CGGACAA 782
DB 701 CGAGGAGCCCGAGGTTGATTGA CTTGTGGGACTTGGCGCAGTCCGCCA CTTCA CGGACAA 760
QY 783 GAGCTGAGAGCGTTTCGGGAGAGCTCAAGCACTTCGAGGCGCAAAATCGAGAGACAA 842
DB 761 GAGCTGAGAGCGTTTCGGGAGAGCTCAAGCACTTCGAGGCGCAAAATCGAGAGACAA 820
QY 843 CCACTACCAAGACAGCTGAGAGATTGCGACAGAGAGCTGAGGACGACAGAGAGCTGGG 902
DB 821 CCACTACCAAGACAGCTGAGAGATTGCGACAGAGAGCTGAGGACGACAGAGAGCTGGG 880
QY 903 CGACGCGAGCGTGTGAGCGCAGCGCGAGAGAGCAAGCCCTGCTCGAGGGGCGGACCAA 962
DB 881 CGACGCGAGCGTGTGAGCGCAGCGCGAGAGAGCAAGCCCTGCTCGAGGGGCGGACCAA 940
QY 963 GAGCTGGGCTACACGGTGAAGAGATCTGAGAGAGCTGTCCGCGAGAGATCTCCAGAGC 1022
DB 941 GAGCTGGGCTACACGGTGAAGAGATCTGAGAGAGCTGTCCGCGAGAGATCTCCAGAGC 1000
QY 1023 TCGCGCGGAGGCA 1035
DB 1001 TCGCGCACAGCA 1013
```

```
RESULT 11
CR596948
LOCUS
DEFINITION
full-length cDNA clone CS0DI035YA24 of Placenta Cot 25-normalized
of Homo sapiens (human).
CR596948
VERSION
CR596948.1 GI:50477755
KEYWORDS
HTC; CNSLT CDNA.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1399)
AUTHORS
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE
Full-length cDNA libraries and normalization
JOURNAL
Unpublished
REMARK
Contact: Feng Liang Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
Genoscope.
2 (bases 1 to 1399)
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY Cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
Location/Qualifiers
1..1399
/molecule="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DI035YA24"
/tissue_type="Placenta Cot 25-normalized"
/plasmid="pCMVSPORT_6"
ORIGIN
```

```
Query Match      25.9%; Score 957; DB 3; Length 1399;
Best Local Similarity 99.0%; Pred. No. 1.2e-185;
Matches 963; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 63 CGGATCTACTCGCGGAGAGAACACGACCCCAAGCGTCCCGCAAAACGGAGTCCGGAGA 122
DB 88 CGGCAAGTACTCGCGGAGAGAACACGACCCCAAGCGTCCCGCAAAACGGAGTCCGGAGA 147
QY 123 GGAGTTCCGATCGAGAGTTGAACACAGCTGTGGGAGAGAGCCCGACGACTGCATCTTCC 182
DB 148 GGAGTTCCGATCGAGAGTTGAACACAGCTGTGGGAGAGAGCCCGACGACTGCATCTTCC 207
QY 183 TCCCGTAGGCTGCCGAGCTCCACGCTGATCTGAAGATACAGGAGAGGAGCGAACTCAT 242
DB 208 TCCCGTAGGCTGCCGAGCTCCACGCTGATCTGAAGATACAGGAGAGGAGCGAACTCAT 267
QY 243 CTGGAAGAACTAAAGCTTGACCGCTTGGAACGAGATGGGAGAGAGGAGGAGGAGTCTAT 302
DB 268 CTGGAAGAACTAAAGCTTGACCGCTTGGAACGAGATGGGAGAGAGGAGGAGGAGTCTAT 327
QY 303 ACGCAACCTCAATGTCTATCTTGGCCCAAGTATGTCTGGAAGAGAGAGGAGCGTCCGCA 362
DB 328 ACGCAACCTCAATGTCTATCTTGGCCCAAGTATGTCTGGAAGAGAGAGGAGCGTCCGCA 387
QY 363 GGTGACCAAGCACTCCCTCAGTGGCAACCCAGAGAGCGGGCTGGATGACCCCGAGCTGA 422
DB 388 GGTGACCAAGCACTCCCTCAGTGGCAACCCAGAGAGCGGGCTGGATGACCCCGAGCTGA 447
QY 423 AAGCTGTGGCAAGCGGAGAGCTCTGGGAAATTTCTCGGCGAGAGAACTGGAGCAAGCT 482
DB 448 AAGCTGTGGCAAGCGGAGAGCTCTGGGAAATTTCTCGGCGAGAGAACTGGAGCAAGCT 507
```

```
QY 483 CTGGCGGAGTTCTCTGATCAAAAAGAGAAAGTTTCAGAGTACAAAGTCTCTCTGTGGAGAC 542
Db 508 CTGGCGGAGTTCTCTGATCACAAGAGAGAAAGTTTCAGAGTACAAAGTCTCTCTGTGGAGAC 567
QY 543 CTTGACGAGCAGCAGGAGAAATCCACAGAGACGTCATTAGCCCTCGGACTCGAGCCACAT 602
Db 568 CTTGACGAGCAGCAGGAGAAATCCACAGAGACGTCATTAGCCCTCGGACTCGAGCCACAT 627
QY 603 CAAAGGCGAGCGTCTCTGACAGCAGGACACACGAGCTCAAGGAGAAAGTCTGCGAGCATCAA 662
Db 628 CAAAGGCGAGCGTCTGACAGCAGGACACACGAGCTCAAGGAGAAAGTCTGCGAGCATCAA 687
QY 663 CAAGGCGCTGAGACCGCTGCGCAGGCTCAGCCACCAAGGCTACAGCACTGAGGCTGAGTT 722
Db 688 CAAGGCGCTGAGACCGCTGCGCAGGCTCAGCCACCAAGGCTCAGCACTGAGGCTGAGTT 747
QY 723 CGAGGAGCCAGGCTGATTGACCTGTGGGACCTGCGGCGAGCTTGGCGGAGCTTGGCGGACAA 782
Db 748 CGAGGAGCCAGGCTGATTGACCTGTGGGACCTTGGCGGAGCTTGGCGGAGCTTGGCGGACAA 807
QY 783 GGAGCTTGGAGGCGCTTCCGGGAGGAGCTCAAGCACTTCAAGGCCAAATCGAGAAGCACAA 842
Db 808 GGAGCTTGGAGGCGCTTCCGGGAGGAGCTCAAGCACTTCAAGGCCAAATCGAGAAGCACAA 867
QY 843 CCACTACCAAGACAGCTGAGATTTGGCAGCAGAAAGCTGAGGCA CGCAGAGAGCGTGGG 902
Db 868 CCACTACCAAGACAGCTGAGATTTGGCAGCAGAAAGCTGAGGCA CGCAGAGAGCGTGGG 927
QY 903 CGACGCGGAGCGTGTGAGCGGACCGCGAGAGCAGCCCTGCTCGAGGGGGGAGCCAA 962
Db 928 CGACGCGGAGCGTGTGAGCGGACCGCGAGAGCAGCCCTGCTCGAGGGGGGAGCCAA 987
QY 963 GGAGCTTGGGCTACACGCTGAAGAGCATCTGACGAGCAGCTGTCCGGGAGGATCTCCAGAGC 1022
Db 988 GGAGCTTGGGCTACACGCTGAAGAGCATCTGACGAGCAGCTGTCCGGGAGGATCTCCAGAGC 1047
QY 1023 TCGCGCGGAGGCA 1035
Db 1048 TCGGCACACGAA 1060

RESULT 12
CR625424
LOCUS
DEFINITION
    1399 bp mRNA linear HTC 21-JUL-2004
full-length cDNA clone CS0DJ007YP06 of T cells (Jurkat cell line)
ACCESSION
CR625424
VERSION
CR625424.1 GI:50506231
KEYWORDS
HTC; CNSLT cDNA.
SOURCE
Homo sapiens
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1399)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact: Feng Liang Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
2 (bases 1 to 1399)
Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail: seque@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
COMMENT
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
Location/Qualifiers
1. .1399
FEATURES
source
```

```
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DJ007YP06"
/tissue_type="T cells (Jurkat cell line) Cot
10-normalized"
/plasmid="pCMVSPORT_6"
```

ORIGIN

```
Query Match 25.9%; Score 957; DB 3; Length 1399;
Best Local Similarity 99.0%; Pred. No. 1.2e-185;
Matches 963; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
```

```
QY 63 CGGATCTCTACGCGGAGAGAAACAGCCCAAGCGCTCCCGAAACGCGAGTCCGGAGA 122
Db 92 CGGCAAGTACTCGCGGAGAGAAACAGCCCAAGCGCTCCCGAAACGCGAGTCCGGAGA 151
QY 123 GGAGTTCGCGATGGAGAGTTGAACACAGCTGTGGGAGAGGCCCGACGACTCATCTTCC 182
Db 152 GGAGTTCGCGATGGAGAGTTGAACAGCTGTGGGAGAGGCCCGACGACTCATCTTCC 211
QY 183 TCCCGTGAAGGCTGGCGGAGCTCCAGCTGATCTGAAGATACAGGAGAGGAGCGAACTCGC 242
Db 212 TCCCGTGAAGGCTGGCGGAGCTCCAGCTGATCTGAAGATACAGGAGAGGAGCGAACTCGC 271
QY 243 CTGGAAGAAACTAAAGCTTTGACGGTTTGA CGAAGATGGGGAGAGGAAAGCGAGCTCAT 302
Db 272 CTGGAAGAAACTAAAGCTTTGACGGTTTGA CGAAGATGGGGAGAGGAAAGCGAGCTCAT 331
QY 303 AGCAACCTCATGTCTTGGCCAGTATGGTCTGTGGACGGAAGAAAGAGCGCTCGGCA 362
Db 332 AGCAACCTCATGTCTTGGCCAGTATGGTCTGTGGACGGAAGAAAGAGCGCTCGGCA 391
QY 363 GGTGACCAACAACTCCCTCAGTGGCACCCAGGAAGACGGGCTGGATGACCCAGGCTGA 422
Db 392 GGTGACCAACAACTCCCTCAGTGGCACCCAGGAAGACGGGCTGGATGACCCAGGCTGA 451
QY 423 AAAGCTGTGGCA CAAAGGCGAAGACCTCTGGGAAATTTCTCCGCGAAGAACTTGACAAGCT 482
Db 452 AAAGCTGTGGCA CAAAGGCGAAGACCTCTGGGAAATTTCTCCGCGAAGAACTTGACAAGCT 511
QY 483 CTGGCGGAGTTCTCTGCATCAAAAGAGAAAGTTCAAGATACAGTACAGTCTCTGTGGAGAC 542
Db 512 CTGGCGGAGTTCTCTGCATCAAAAGAGAAAGTTCAAGATACAGTACAGTCTCTGTGGAGAC 571
QY 543 CTTGACGAGCAGCAGGAGAAATCCACAGAAAGTCTATTAGCCCTCGGACTCGAGCGACAT 602
Db 572 CTTGACGAGCAGCAGGAGAAATCCACAGAAAGTCTATTAGCCCTCGGACTCGAGCGACAT 631
QY 603 CAAAGGCGAGCGTCTCTGACAGCAGGACAC CCGAGCTGAAGAGGAGAGTCTGCGAGCATCAA 662
Db 632 CAAAGGCGAGCGTCTCTGACAGCAGGACAC CCGAGCTGAAGAGGAGAGTCTGCGAGCATCAA 691
QY 663 CAAGGCGCTGAGACCGCTCGCGAGGCTAGCCACCAAGGCTACAGCACTGAGGCTGAGTT 722
Db 692 CAAGGCGCTGAGACCGCTCGCGAGGCTAGCCACCAAGGCTACAGCACTGAGGCTGAGTT 751
QY 723 CGAGGAGCCAGGCTGATTGACCTGTGGGACCTTGGGCGAGTCCGCGAGTCTCACGAGCAA 782
Db 752 CGAGGAGCCAGGCTGATTGACCTGTGGGACCTTGGGCGAGTCCGCGAGTCTCACGAGCAA 811
QY 783 GGAGCTTGAAGGCGTTCCGGGAGGAGCTCAAGCACTTCAAGGCCAAATCGAGAAGCACAA 842
Db 812 GGAGCTTGAAGGCGTTCCGGGAGGAGCTCAAGCACTTCAAGGCCAAATCGAGAAGCACAA 871
QY 843 CCACTACCAAGACAGCTGAGATTTGGCAGCAGAAAGCTGAGGACCGCAGAGAGCGTGGG 902
Db 872 CCACTACCAAGACAGCTGAGATTTGGCAGCAGAAAGCTGAGGACCGCAGAGAGCGTGGG 931
QY 903 CGACGCGGAGCTGTGAGCGGCGAGAGGAGCAGCCCTGCTGTGGAGGGGGAGCCAA 962
Db 932 CGACGCGGAGCTGTGAGCGGCGAGAGGAGCAGCCCTGCTGTGGAGGGGGAGCCAA 991
```

```
QY 963 GGAGCTGGGCTACACGGTGAAGAGCATCTGCAGGACCTGTCCGGCAGGATCTCCAGAGC 1022
DB 992 GGAGCTGGGCTACACGGTGAAGAGCATCTGCAGGACCTGTCCGGCAGGATCTCCAGAGC 1051
QY 1023 TCGCGCCGAGGCA 1035
DB 1052 TCGGCACAACGAA 1064

RESULT 13
CR591537
LOCUS
DEFINITION
Full-length cDNA clone CS0D005YL06 of T cells (Jurkat cell line)
Cot 10-normalized of Homo sapiens (human).
CR591537
ACCESSION
VERSION
CR591537.1 GI:50472344
KEYWORDS
HTC; CNSLT_cDNA.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1402)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
JOURNAL
REMARK
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
2 (bases 1 to 1402)
Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
source
1. .1402
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D005YL06"
/tissue_type="T cells (Jurkat cell line) Cot
10-normalized"
/plasmid="pCMVSPORT_6"

ORIGIN
Query Match 25.9%; Score 957; DB 3; Length 1402;
Best Local Similarity 99.0%; Pred. No. 1.2e-185;
Matches 963; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY 63 CGGATCCTTACTCGCGGAGAGAAACACAGCCCAAGCGTCCCGAAACGGAGTCCGGAGA 122
DB 76 CGGCAAGTACTCGCGGAGAGAAACACAGCCCAAGCGTCCCGAAACGGAGTCCGGAGA 135
QY 123 GGAGTTCCCGATGGAGAGTTGAACACGCTGTGGGAGAGGCCACCGACTGCATCTTCC 182
DB 136 GGAGTTCCCGATGGAGAGTTGAACACGCTGTGGGAGAGGCCACCGACTGCATCTTCC 195
QY 183 TCCGCTGAGGCTGGCCGAGCTCCACGCTGATCTGAAGATACAGGAGGAGCGAACTCCG 242
DB 196 TCCGCTGAGGCTGGCCGAGCTCCACGCTGATCTGAAGATACAGGAGGAGCGAACTCCG 255
QY 243 CTGGAGAGAACTAAAGCTTGAAGCTTGAAGATGGGGAGAGAAAGCGAGACTCAT 302
DB 256 CTGGAGAGAACTAAAGCTTGAAGCTTGAAGATGGGGAGAGAAAGCGAGACTCAT 315
QY 303 ACGCAACCTCAATGTCTATCTGCGCAAGTATGCTGTGGACGGAAGAAAGCGCTCGCA 362
DB 316 ACGCAACCTCAATGTCTATCTGCGCAAGTATGCTGTGGACGGAAGAAAGCGCTCGCA 375
```

```
QY 363 GGTGACCAAGCAACTCCCTCAGTGGCACCCAGGAAGACGGGCTGGATGATCCCCAGGCTGA 422
DB 376 GGTGACCAAGCAACTCCCTCAGTGGCACCCAGGAAGACGGGCTGGATGATCCCCAGGCTGA 435
QY 423 AAAGCTGTGGCACAAAGGCGAAGACCTCTGGGAAATTTCTCCGGGGAAGAACTGGACAGCT 482
DB 436 AAAGCTGTGGCACAAAGGCGAAGACCTCTGGGAAATTTCTCCGGGGAAGAACTGGACAGCT 495
QY 483 CTGGCGGGAGTTCTCTGCATCAAAAGAGAAAGTTCACGAGTACAAAGCTCTCTCTGGAGAC 542
DB 496 CTGGCGGGAGTTCTCTGCATCAAAAGAGAAAGTTCACGAGTACAAAGCTCTCTCTGGAGAC 555
QY 543 CTTGACGAGGACCGAAGAAATTCACGAGAAACGTCTTAGCCCTCGGACCTGAGCGACAT 602
DB 556 CTTGACGAGGACCGAAGAAATTCACGAGAAACGTCTTAGCCCTCGGACCTGAGCGACAT 615
QY 603 CAAAGGCGAGCTCTCTGCACAGCAGGACACACGAGAGCTGAAGAGAAAGCTGCGAGCATCAA 662
DB 616 CAAAGGCGAGCTCTCTGCACAGCAGGACACACGAGAGCTGAAGAGAAAGCTGCGAGCATCAA 675
QY 663 CCAGGCGCTTGGACCCGCTCGCAGGCTCAGCCACGAGGCTACAGCACTGAGGCTGAGTT 722
DB 676 CCAGGCGCTTGGACCCGCTCGCAGGCTCAGCCACGAGGCTACAGCACTGAGGCTGAGTT 735
QY 723 CGAGGAGCCCGAGGCTGATTGACCTCTGGGACCTCGGCGAGTCCGCCAACCTCACGACAA 782
DB 736 CGAGGAGCCCGAGGCTGATTGACCTCTGGGACCTCGGCGAGTCCGCCAACCTCACGACAA 795
QY 783 GGAGCTGGAGGCGTTCCGGGAGGAGCTCAGCACTTCGAGCAAAATTCGAGAAGCACAA 842
DB 796 GGAGCTGGAGGCGTTCCGGGAGGAGCTCAGCACTTCGAGCAAAATTCGAGAAGCACAA 855
QY 843 CCACCTACCAAGACAGCTGGAGATTGCGCACGAGAAAGCTGAGGACGACGAGAGCGTGGG 902
DB 856 CCACCTACCAAGACAGCTGGAGATTGCGCACGAGAAAGCTGAGGACGACGAGAGCGTGGG 915
QY 903 CGACGCGAGCGTGTGAGCCGCGAGAGCAACGCTCTGTGGAGGGGCGGACCAAA 962
DB 916 CGACGCGAGCGTGTGAGCCGCGAGAGCAACGCTCTGTGGAGGGGCGGACCAAA 975
QY 963 GGAGCTGGGCTACACGGTGAAGAGCATCTGCAGGACCTGTCCGCGAGGATCTCCAGAGC 1022
DB 976 GGAGCTGGGCTACACGGTGAAGAGCATCTGCAGGACCTGTCCGCGAGGATCTCCAGAGC 1035
QY 1023 TCGCGCCGAGGCA 1035
DB 1036 TCGGCACAACGAA 1048

RESULT 14
CR619887
LOCUS
DEFINITION
full-length cDNA clone CS0DE003YG19 of Placenta of Homo sapiens
(human).
CR619887
ACCESSION
VERSION
CR619887.1 GI:50500694
KEYWORDS
HTC; CNSLT_cDNA.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1409)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
JOURNAL
REMARK
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
2 (bases 1 to 1409)
Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
```

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.

FEATURES
source Location/Qualifiers
1..1409
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D003YB19"
/tissue_type="Placenta"
/plasmid="pCMVSPORT_6"

ORIGIN
Query Match 25.9%; Score 957; DB 3; Length 1409;
Best Local Similarity 99.0%; Pred. No. 1.2e-185;
Matches 963; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 63 CGGATCTTACTCCGGGAGAGAACAGCCCAAGCCGTCGCCGAACGGAGTCCGGAGA 122
DB |||||
QY 92 CGCAAGTACTCCGGGAGAGAACAGCCCAAGCCGTCGCCGAACGGAGTCCGGAGA 151
DB |||||
QY 123 GGAGTTCCGCATCGAAGTTGAACAGCTGTGGGAGAGGCCAGCGACTGCTTCC 182
DB |||||
QY 152 GGAGTTCCGCATCGAAGTTGAACAGCTGTGGGAGAGGCCAGCGACTGCTTCC 211
DB |||||
QY 183 TCCGTTGAGGCTGCCAGCTCCAGCTGATCTGAAGATACAGGAGGGAGCAATCCG 242
DB |||||
QY 212 TCCGTTGAGGCTGCCAGCTCCAGCTGATCTGAAGATACAGGAGGGAGCAATCCG 271
DB |||||
QY 243 CTGGAAGAACTAAAGCTTGACGGCTTGACGAAGATGGGAGAAAGAGCAAGCACTCAT 302
DB |||||
QY 272 CTGGAAGAACTAAAGCTTGACGGCTTGACGAAGATGGGAGAAAGAGCAAGCACTCAT 331
DB |||||
QY 303 AGCAACCTCAATGTCTATTTGCCAAGTATGTCTTGGACGGAAAGAGCAAGCTCGGCA 362
DB |||||
QY 332 AGCAACCTCAATGTCTATTTGCCAAGTATGTCTTGGACGGAAAGAGCAAGCTCGGCA 391
DB |||||
QY 363 GGTGACAGCACTCCTCAGTGGCCAGCCAGGAGAGCGGCTGGATGATCCCGAGGTGA 422
DB |||||
QY 392 GGTGACAGCACTCCTCAGTGGCCAGCCAGGAGAGCGGCTGGATGATCCCGAGGTGA 451
DB |||||
QY 423 AAAGCTGTGSCACAGCGAAGACCTCTGGGAAATTTCTCGGCGAAGAACTGGACAAGCT 482
DB |||||
QY 452 AAAGCTGTGSCACAGCGGAGACCTCTGGGAAATTTCTCGGCGAAGAACTGGACAAGCT 511
DB |||||
QY 483 CTGGCGGAGTTCTCTCATCACAAAGAGAAAGTTTACAGATCAACGTCCTGCTGGAGAC 542
DB |||||
QY 512 CTGGCGGAGTTCTCTCATCACAAAGAGAAAGTTTACAGATCAACGTCCTGCTGGAGAC 571
DB |||||
QY 543 CTTGACAGGACCGAGAAATCAACAGACGTCATTAGCCCTCCGACCTGAGCCACAT 602
DB |||||
QY 572 CTTGACAGGACCGAGAAATCAACAGACGTCATTAGCCCTCCGACCTGAGCCACAT 631
DB |||||
QY 603 CAAGGCGAGCTCTGCACAGCAGGACACAGGAGCTGAAGGAGAAAGCTGGCAGCATCAA 662
DB |||||
QY 632 CAAGGCGAGCTCTGCACAGCAGGACACAGGAGCTGAAGGAGAAAGCTGGCAGCATCAA 691
DB |||||
QY 663 CCAAGGCTTGGACCGCTCGCAGGAGTCAAGCAACAGGAGTACAGCACTGAGGCTGAGTT 722
DB |||||
QY 692 CCAAGGCTTGGACCGCTCGCAGGAGTCAAGCAACAGGAGTACAGCACTGAGGCTGAGTT 751
DB |||||
QY 723 CGAGGAGCCAGGAGTATGACCTGTGGACCTGGCGAGTCCGCCAACCTTACGAGCAA 782
DB |||||
QY 752 CGAGGAGCCAGGAGTATGACCTGTGGACCTGGCGAGTCCGCCAACCTTACGAGCAA 811
DB |||||
QY 783 GGAGCTGAGGAGCTTCCGGGAGGAGCTCAAGCACTTTCGAAGCCCAATCGAGAGCACAA 842
DB |||||
QY 812 GGAGCTGAGGAGCTTCCGGGAGGAGCTCAAGCACTTTCGAAGCCCAATCGAGAGCACAA 871
DB |||||

QY 843 CCACTACCAAGACAGCTCGAGATTCCGACAGAAAGCTTGAGCACGACGACAGAGCGCTGGG 902
DB |||||
QY 872 CCACTACCAAGACAGCTCGAGATTCCGACAGAAAGCTTGAGCACGACGACAGAGCGCTGGG 931
DB |||||
QY 903 CGACGCGAGCTGTGAGCCGAGCCGAGAGGACGACGCTGCTGAGGGGGCGGACCAA 962
DB |||||
QY 932 CGACGCGAGCTGTGAGCCGAGCCGAGAGGACGACGCTGCTGAGGGGGCGGACCAA 991
DB |||||
QY 963 GGAGCTGGGCTACACGGTGAAGAAGCATCTCGAGGACCTGTCCGGCAGGATCTCCAGAGC 1022
DB |||||
QY 992 GGAGCTGGGCTACACGGTGAAGAAGCATCTCGAGGACCTGTCCGGCAGGATCTCCAGAGC 1051
DB |||||
QY 1023 TCGCGCCGAGGCA 1035
DB |||||
QY 1052 TCGGCACAAACGAA 1064
DB |||||

RESULT 15
CR607197 1411 bp mRNA linear HTC 21-JUL-2004
LOCUS full-length cDNA clone CS0D1023YB19 of Placenta Cot 25-normalized
DEFINITION of Homo sapiens (human).
ACCESSION CR607197
VERSION CR607197.1 GI:50488004
KEYWORDS HTC; CNSLT_CDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1411)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normaliza
JOURNAL Unpublished
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paraday Avenue
2 (bases 1 to 1411)
REFERENCE Genoscope.
AUTHORS Direct Submission
TITLE Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.

FEATURES
source Location/Qualifiers
1..1411
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1023YB19"
/tissue_type="Placenta Cot 25-normalized"
/plasmid="pCMVSPORT_6"

ORIGIN
Query Match 25.9%; Score 957; DB 3; Length 1411;
Best Local Similarity 99.0%; Pred. No. 1.2e-185;
Matches 963; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 63 CGGATCTTACTCCGGGAGAGAACAGCCCAAGCCGTCGCCGAACGGAGTCCGGAGA 122
DB |||||
QY 92 CGCAAGTACTCCGGGAGAGAACAGCCCAAGCCGTCGCCGAACGGAGTCCGGAGA 149
DB |||||
QY 123 GGAGTTCCGCATCGAAGTTGAACAGCTGTGGGAGAGGCCAGCGACTGCTTCC 182
DB |||||
QY 150 GGAGTTCCGCATCGAAGTTGAACAGCTGTGGGAGAGGCCAGCGACTGCTTCC 209
DB |||||
QY 183 TCCGTTGAGGCTGCCAGCTCCAGCTGATCTGAAGATACAGGAGGGAGCAACTCCG 242
DB |||||
QY 210 TCCGTTGAGGCTGCCAGCTCCAGCTGATCTGAAGATACAGGAGGGAGCAACTCCG 269
DB |||||

```
QY 243 CTGGAAGAACTAAAGCTTGA CGGCTTGA CGAAGATGGGAGAGAGAAAGCGAGACTCAT 302
Db      |||
QY 270 CTGGAAGAACTAAAGCTTGA CGGCTTGA CGAAGATGGGAGAGAGAAAGCGAGACTCAT 329
Db      |||
QY 303 AGCAACCTCNAATGTCATCTTGGCCCAAGTATGGTCTGACGCGAAGAAAGACGCTCGGCA 362
Db      |||
QY 330 AGCAACCTCNAATGTCATCTTGGCCCAAGTATGGTCTGACGCGAAGAAAGACGCTCGGCA 389
Db      |||
QY 363 GGTGACCAAGCACTCCCTCAGTGGCAACCAGGAAGACGGGCTGGATGACCCAGGCTGGA 422
Db      |||
QY 390 GGTGACCAAGCACTCCCTCAGTGGCAACCAGGAAGACGGGCTGGATGACCCAGGCTGGA 449
Db      |||
QY 423 AAAGCTGTGGCACAAGCGCAAGACCTCTGGGAAATTTCTCCGCGAAGAACTGGACAAAGCT 482
Db      |||
QY 450 AAAGCTGTGGCACAAGCGCAAGACCTCTGGGAAATTTCTCCGCGAAGAACTGGACAAAGCT 509
Db      |||
QY 483 CTGGCGGGAGTTCTTCGTCATCAAAAGAAAGTTTACGAGTACAAAGTCTCTGCTGGAGAC 542
Db      |||
QY 510 CTGGCGGGAGTTCTTCGTCATCAAAAGAAAGTTTACGAGTACAAAGTCTCTGCTGGAGAC 569
Db      |||
QY 543 CTTGAGCAGGACCGAAGAAATCCACGAGAACGTCATTAGCCCTCGGACCTGAGCGACAT 602
Db      |||
QY 570 CTTGAGCAGGACCGAAGAAATCCACGAGAACGTCATTAGCCCTCGGACCTGAGCGACAT 629
Db      |||
QY 603 CAAAGGCGAGCTCTTGCACAGCAGGCAACGCGAGCTGAAAGGAGAACTGCGCAGCATCAA 662
Db      |||
QY 630 CAAAGGCGAGCTCTTGCACAGCAGGCAACGCGAGCTGAAAGGAGAACTGCGCAGCATCAA 689
Db      |||
QY 663 CCAGGCGCTTGGACCGGCTCGGACAGGCTCAGCCACAGGGCTACAGCACTGAGGCTGAGTT 722
Db      |||
QY 690 CCAGGCGCTTGGACCGGCTCGGACAGGCTCAGCCACAGGGCTACAGCACTGAGGCTGAGTT 749
Db      |||
QY 723 CGAGGAGCCAGGGTGATTGACCTGTGGGACCTTGGCGCAGTCCGCCAACCTCACCGACAA 782
Db      |||
QY 750 CGAGGAGCCAGGGTGATTGACCTGTGGGACCTTGGCGCAGTCCGCCAACCTCACCGACAA 809
Db      |||
QY 783 GGAGCTGGAGGGCTTCCGGAGGAGCTCAAGCACTTCGAAGCCAAATCGAAGACACAA 842
Db      |||
QY 810 GGAGCTGGAGGGCTTCCGGAGGAGCTCAAGCACTTCGAAGCCAAATCGAAGACACAA 869
Db      |||
QY 843 CCACTACCAAGACGAGCTGGAGNTTCCGCAAGAGCTGAGGCAAGAGAGGCTGGG 902
Db      |||
QY 870 CCACTACCAAGACGAGCTGGAGNTTCCGCAAGAGCTGAGGCAAGAGAGGCTGGG 929
Db      |||
QY 903 CGACGCGAGCGTGTGAGCCGAGCGCGAGAGCAACGCCCTGCTGGAGGGCGGACCAA 962
Db      |||
QY 930 CGACGCGAGCGTGTGAGCCGAGCGCGAGAGCAACGCCCTGCTGGAGGGCGGACCAA 989
Db      |||
QY 963 GGAGCTGGGCTACACGGTGAAGAGCATCTGAGGACCTGTCCGGCAGGATTTCCAGAGC 1022
Db      |||
QY 990 GGAGCTGGGCTACACGGTGAAGAGCATCTGAGGACCTGTCCGGCAGGATTTCCAGAGC 1049
Db      |||
QY 1023 TCGCGCCGAGGCA 1035
Db      |||
QY 1050 TCGGCAACGAA 1062
Db      |||
```

Search completed: April 4, 2005, 18:57:30
Job time: 10494 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2005, 16:02:35 ; Search time 132 Seconds
(without alignments)
3598.044 Million cell updates/sec

Title: US-10-600-862A-7
Perfect score: 6522
Sequence: 1 MRGPGALWLLALRTVLGS.....KVLDICVSLLMGEQFLVSWC 1228

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*

1: Geneseq1980s:*

2: Geneseq1990s:*

3: Geneseq2000s:*

4: Geneseq2001s:*

5: Geneseq2002s:*

6: Geneseq2003as:*

7: Geneseq2003bs:*

8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4744.5	72.7	952	8 ADQ91896	Adq91896 Human glu
2	4744.5	72.7	953	7 ADD45733	Add45733 Human Pro
3	4741	72.7	883	7 ABM79002	Abm79002 Human alp
4	4724.5	72.4	950	8 ADF47521	Adf47521 Human GAA
5	4283.5	65.7	878	8 ABM84315	Abm84315 Human dia
6	4260.5	65.3	904	8 ABM84314	Abm84314 Human dia
7	3974	60.9	953	5 ABB57174	Abb57174 Mouse isc
8	3763.5	57.7	788	8 ABM84316	Abm84316 Human dia
9	2024	31.0	707	7 ADJ83134	Adj83134 Glucosyl
10	2007.5	30.8	638	8 ABM84317	Abm84317 Human dia
11	1900.5	29.1	1857	8 ADQ19560	Adq19560 Human PRO
12	1877.5	28.8	1827	2 AAU74090	Aau74090 Human hsi
13	1877.5	28.8	1827	4 AAU09028	Aau09028 Human suc
14	1877.5	28.8	1829	4 ABG14994	Abg14994 Novel hum
15	1660.5	25.5	357	4 AAB48844	Aab48844 Human RAP
16	1660.5	25.5	357	5 AAQ18621	Aaq18621 Human rec
17	1660.5	25.5	357	7 ADD44975	Add44975 Human Pro
18	1660.5	25.5	357	7 ADD44979	Add44979 Human Pro
19	1660.5	25.5	357	7 ADD44971	Add44971 Human Pro
20	1660.5	25.5	357	7 ADD44967	Add44967 Human Pro
21	1660.5	25.5	357	7 ADD44963	Add44963 Human Pro
22	1660.5	25.5	357	7 ADD44959	Add44959 Human Pro
23	1660.5	25.5	357	7 ADD44952	Add44952 Human Pro
24	1660.5	25.5	357	7 ADJ75361	Adj75361 Marker ge
25	1660.5	25.5	357	8 ADQ91471	Adq91471 Amino aci

26	1660.5	25.5	357	8 ADR23135	Adp23135 PRO polyp
27	1660.5	25.5	682	8 ADQ39604	Adq39604 Human myo
28	1660.5	25.5	682	8 ADQ39605	Adq39605 Human myo
29	1650	25.3	323	2 AAR56282	Aar56282 Human tis
30	1647	25.3	323	2 AAR80944	Aar80944 Receptor -
31	1498.5	23.0	902	6 ABU63302	Abu63302 Wild-type
32	1472	22.6	901	5 ABB09151	Abb09151 Buckwheat
33	1472	22.6	901	5 AAU97731	Aau97731 Common bu
34	1459.5	22.4	646	8 ADQ67734	Adq67734 Novel hum
35	1459.5	22.4	903	6 ABU63301	Abu63301 Wild-type
36	1457	22.3	901	5 ABB09152	Abb09152 Buckwheat
37	1457	22.3	901	5 AAU97732	Aau97732 Common bu
38	1403.5	21.5	877	2 AAW59040	Aaw59040 Barley al
39	1401.5	21.5	914	6 ABU63150	Abu63150 Wild-type
40	1392.5	21.4	969	7 ADC51564	Adc51564 Schizosac
41	1392.5	21.4	969	8 ADH61320	Adh61320 Schizosac
42	1387	21.3	877	6 ABU63305	Abu63305 Barley al
43	1381	21.2	877	6 ABU63304	Abu63304 Barley al
44	1380	21.2	877	6 ABU63308	Abu63308 Barley al
45	1380	21.2	877	6 ABU63307	Abu63307 Barley al

ALIGNMENTS

RESULT 1
ADQ91896
ID ADQ91896 standard; protein; 952 AA.
XX
AC ADQ91896;
XX
DT 21-OCT-2004 (first entry)
XX
DE Human glucosidase alpha acid (GAA) protein SEQ ID NO:2.
XX
KW chimeric protein; secretory signal; lysosomal protein;
KW lysosomal acid alpha-glucosidase; glucosidase alpha acid; GAA; enzyme;
KW lysosomal acid alpha-glucosidase deficiency; GAA deficiency;
KW glycogen storage disease type II; GSD II; human; chromosome 17.
XX
OS Homo sapiens.
XX
FN WO2004064750-A2.
XX
PD 05-AUG-2004.
XX
PF 21-JAN-2004; 2004WO-US001453.
XX
PR 22-JAN-2003; 2003US-0441789P.
XX
PA (UYDU-) UNIV DUKE.
XX
PI Koeberl DD, Sun B;
XX
DR WPI; 2004-571599/55.
DR N-PSDB; ADQ91895, ADQ91897.
DR GENBANK; NM_000152.
XX
PT New nucleic acid expressing lysosomal acid alpha-glucosidase (GAA)
PT polypeptide, useful for preparing a composition for treating e.g.,
PT glycogen storage disease type II.
XX
PS Example 13; SEQ ID NO 2; 128pp; English.
XX
CC The present invention describes an isolated nucleic acid (I) encoding a
CC chimeric polypeptide comprising a secretory signal sequence operably
CC linked to a lysosomal polypeptide. Also described: (1) a vector
CC comprising the isolated nucleic acid; (2) a pharmaceutical formulation
CC comprising the isolated nucleic acid in a carrier; (3) a cell comprising
CC the isolated nucleic acid; (4) a chimeric polypeptide comprising a
CC secretory signal sequence operably linked to a lysosomal polypeptide; (5)
CC delivering a nucleic acid encoding a lysosomal polypeptide (preferably a
CC lysosomal acid alpha-glucosidase (GAA)) to a cell; (6) producing a GAA

method for identifying a compound or small molecule that regulates the activity in an animal of one or more of the polypeptides given in the specification, a method for identifying a compound useful in treating pain and a pharmaceutical composition comprising the one or more polypeptides or their antibodies. The polynucleotide or the compound that modulates its activity is useful for preparing a medicament for treating pain (e.g. spinal segmental nerve injury (Chung), chronic constriction injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene therapy). The sequence presented is a human protein (shown in Table 2 of the specification) which is differentially expressed during pain. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic form directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 953 AA;

Query Match 72.7%; Score 4744.5; DB 7; Length 953;
Best Local Similarity 97.2%; Pred. No. 0;
Matches 890; Conservative 3; Mismatches 18; Indels 5; Gaps 1;

QY 313 LEGRTKELGYVVKXHLQDLSGRISRAAEATGAHPRPRAVPTQCDVPPNSRFDCAADK 372
DB 42 LSGSPVLEETHAQQGASRPGPR-----DAQAHPRPRAVPTQCDVPPNSRFDCAADK 96

QY 373 AITQEQCEARGCCYIPAKQLOQAQWQPCPPPPSPSYKLENLSSEMGYTATLTRT 432
DB 97 AITQEQCEARGCCYIPAKQLOQAQWQPCPPPPSPSYKLENLSSEMGYTATLTRT 156

QY 433 PTFPPKDIILTRLDVMMETENRHLFTIKDPAHRYEVLPTPRVHRAPSPLYSVEFSEE 492
DB 157 PTFPPKDIILTRLDVMMETENRHLFTIKDPAHRYEVLPTPRVHRAPSPLYSVEFSEE 216

QY 493 PFGVIVHRQLDGRVLLNTTVAFLFFADQLQLSTLSPSQYITGLAEHLSPLMLSTWTRI 552
DB 217 PFGVIVHRQLDGRVLLNTTVAFLFFADQLQLSTLSPSQYITGLAEHLSPLMLSTWTRI 276

QY 553 TLWNRDLATPGANLYGSHPPFYALBEGGSAHGCVLLNSNMDVVLQSPALSWRSTGGI 612
DB 277 TLWNRDLATPGANLYGSHPPFYALBEGGSAHGCVLLNSNMDVVLQSPALSWRSTGGI 336

QY 613 LDVYIFLGPPEKSVQOYLVDVGYPPMPYVWIGLHCRWGSSTAITQVVENMTRAHF 672
DB 337 LDVYIFLGPPEKSVQOYLVDVGYPPMPYVWIGLHCRWGSSTAITQVVENMTRAHF 396

QY 673 PLDVQWNLDDYMSRRDFTFNKDGFRDFPAMVQELHQQGRRYMMIVDPAISSGGPAGSYR 732
DB 397 PLDVQWNLDDYMSRRDFTFNKDGFRDFPAMVQELHQQGRRYMMIVDPAISSGGPAGSYR 456

QY 733 PYDEGLRRGVFTNETGQPLIGKWPFGSTAFDPFTNPTALAWMEDMVAFHDOVPPDGLW 792
DB 457 PYDEGLRRGVFTNETGQPLIGKWPFGSTAFDPFTNPTALAWMEDMVAFHDOVPPDGLW 516

QY 793 IDWNEPSNFRGSEDCPNNELENPPYVGVGGTLOAATICASSHQPLSTHYNLNLXG 852
DB 517 IDWNEPSNFRGSEDCPNNELENPPYVGVGGTLOAATICASSHQPLSTHYNLNLXG 576

QY 853 LTEATASHRALVKARTRFPVISRSTFAGHGRYAGHTGDMVSSWQLASSVPEILQFNL 912
DB 577 LTEATASHRALVKARTRFPVISRSTFAGHGRYAGHTGDMVSSWQLASSVPEILQFNL 636

QY 913 LGVPLVGADVCGFLGNTSBEELCVRTQLGAFYPPFMNHNLSLSLPQEPYSFSEPAQAMR 972
DB 637 LGVPLVGADVCGFLGNTSBEELCVRTQLGAFYPPFMNHNLSLSLPQEPYSFSEPAQAMR 696

QY 973 KALTLYALLPHLYTLFQAHVAGETVARPLFLFEPFKDSSTVDHQLLWGAALLTPVL 1032
DB 697 KALTLYALLPHLYTLFQAHVAGETVARPLFLFEPFKDSSTVDHQLLWGAALLTPVL 756

QY 1033 QAGKAEVTVGFFLGTWYDLQTVIEALGSLPPPPAAPREPAIHSEQWVTLPAPLDTINV 1092
DB 757 QAGKAEVTVGFFLGTWYDLQTVIEALGSLPPPPAAPREPAIHSEQWVTLPAPLDTINV 816

QY 1093 HLRAGYIIPLOQFGLTTTTSRQOPMALAVALTKGGEARGELFWDDGESLEVLERGAYTQV 1152

DB 817 HLRAGYIIPLOQFGLTTTTSRQOPMALAVALTKGGEARGELFWDDGESLEVLERGAYTQV 876
QY 1153 IFLARNNTIVNELVRVTSEGAGLQKQVTLGVATAPQQVLSNGVPSVSNFTYSPDTKVL 1212
DB 877 IFLARNNTIVNELVRVTSEGAGLQKQVTLGVATAPQQVLSNGVPSVSNFTYSPDTKVL 936

QY 1213 ICVSLLMGSEQFLVSWC 1228
DB 937 ICVSLLMGSEQFLVSWC 952

RESULT 3
ABM79002 ID ABM79002 standard; protein; 883 AA.
XX AC ABM79002;
XX DT 15-JAN-2004 (first entry)
XX DE Human alpha-galactosidase (mature polypeptide).
XX KW Human; alpha-glucosidase; lysosome; enzyme; Pompe disease;
XX KM transgenic plant.
XX OS Homo sapiens.
XX PN WO2003073839-A2.
XX PD 12-SEP-2003.
XX PF 03-MAR-2003; 2003WO-IT000120.
XX PR 01-MAR-2002; 2002IT-RM000115.
XX PA (PLAN-) PLANTECHNO SRL.
XX PI Fogher C, Reggi S;
XX DR WPI; 2003-712829/67.
XX N-PSDB; ACF80584.
XX PT New genetically transformed plant that can produce a lysosomal enzyme of animal or human origin, useful for preparing a medicament for enzyme replacement therapy in Gaucher, Anderson-Fabry or Pompe disease.
XX PS Example 10; Page 51-53; 53pp; English.
XX CC The present sequence is the protein sequence of human alpha-glucosidase (GAA). A deficiency of this enzyme causes Pompe disease. The invention is based on the discovery that lysosomal enzymes, such as GAA, can be expressed in seed storage organs in a form which is stable (over 12 months in stored seeds), enzymatically active and in a high amount suitable for medical use. An expression cassette includes the GAA coding sequence (minus the native signal sequence), and basic 7S soy globulin promoter and signal sequences. Such constructs can be used to express human GAA in transgenic plants, especially plants having a high protein content, e.g. legumes, cereals and tobacco

XX SQ Sequence 883 AA;

Query Match 72.7%; Score 4741; DB 7; Length 883;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 882; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 346 AHGPRPRAVPTQCDVPPNSRFDCAADKAITQEQCEARGCCYIPAKQLOQAQWQPCFF 405
DB 1 AHGPRPRAVPTQCDVPPNSRFDCAADKAITQEQCEARGCCYIPAKQLOQAQWQPCFF 60

QY 406 PPSPSYKLENLSSEMGYTATLTRTTPFPKDIILTRLDVMMETENRHLFTIKDPAHRYEVLPTPRVHRAPSPLYSVEFSEE 465
DB 61 PPSPSYKLENLSSEMGYTATLTRTTPFPKDIILTRLDVMMETENRHLFTIKDPAHRYEVLPTPRVHRAPSPLYSVEFSEE 120

```
QY 466 RYEVPLETRVHSHRAPSPLYSVFSEBEPFGVIVHRLDGRVLLNTTVAELFPADQFLQLS 525
Db 121 RYEVPLETRVHSHRAPSPLYSVFSEBEPFGVIVHRLDGRVLLNTTVAELFPADQFLQLS 180
QY 526 TSLPSQYITGLAEHLSPMLSTSWTRITLWNRDLAPTGCANLYGSHPPFYLALEDGGSAGH 585
Db 181 TSLPSQYITGLAEHLSPMLSTSWTRITLWNRDLAPTGCANLYGSHPPFYLALEDGGSAGH 240
QY 586 VFLLNSNAMDVVLQSPALSWRSTGILDVYIFLGPPEKSVVQQYLDVVGVYFPMPPYWG 645
Db 241 VFLLNSNAMDVVLQSPALSWRSTGILDVYIFLGPPEKSVVQQYLDVVGVYFPMPPYWG 300
QY 646 GFHLCKWGSSTAITQVVENMTRAHPPDLVDQNDLDYMDSRDFTFNKDGGRDPPAMVQ 705
Db 301 GFHLCKWGSSTAITQVVENMTRAHPPDLVDQNDLDYMDSRDFTFNKDGGRDPPAMVQ 360
QY 706 ELHQGGRYMMIVDPAISSGPGAGSRPYDEGLRRCGVETNETGQPLIGKVPWGSTAFPD 765
Db 361 ELHQGGRYMMIVDPAISSGPGAGSRPYDEGLRRCGVETNETGQPLIGKVPWGSTAFPD 420
QY 766 FNTPTALAWMEDVAFHFQVDPDGLWIDWNEPSNFIKSGDCPNNELENPPYVGVVG 825
Db 421 FNTPTALAWMEDVAFHFQVDPDGLWIDWNEPSNFIKSGDCPNNELENPPYVGVVG 480
QY 826 GTLQAAATICASSHQFTSTHYNLHNLVGLTEATASHRALVKARTRPFVSRSTFAGHGXY 885
Db 481 GTLQAAATICASSHQFTSTHYNLHNLVGLTEATASHRALVKARTRPFVSRSTFAGHGXY 540
QY 886 AGHWTGDVWSSWELQASSVPEILQFNLLGVPLVGDVCGFLGNTSEELCVRTQLGAFYP 945
Db 541 AGHWTGDVWSSWELQASSVPEILQFNLLGVPLVGDVCGFLGNTSEELCVRTQLGAFYP 600
QY 946 FMENHNSLLSLPOEYFSFSEPAQAMRKALTRYALLPHLYTLFHQAHVAGETVARPLFL 1005
Db 601 FMENHNSLLSLPOEYFSFSEPAQAMRKALTRYALLPHLYTLFHQAHVAGETVARPLFL 660
QY 1006 EFPKDSSTWVDHQLWGBALLITPVLOAKGAEVTGYFFLGTWYDQLQVPIEALGSLPPP 1065
Db 661 EFPKDSSTWVDHQLWGBALLITPVLOAKGAEVTGYFFLGTWYDQLQVPIEALGSLPPP 720
QY 1066 PAAPREPAHSEGWTLPAFLDTINVHLRAGVILPLOGPGLTTTSSROQPWALAVALK 1125
Db 721 PAAPREPAHSEGWTLPAFLDTINVHLRAGVILPLOGPGLTTTSSROQPWALAVALK 780
QY 1126 GGEARGELFWDDGESLEVLERGAYTVIFLARNTTVNELVRVTSGAGLQKQVTLGV 1185
Db 781 GGEARGELFWDDGESLEVLERGAYTVIFLARNTTVNELVRVTSGAGLQKQVTLGV 840
QY 1186 ATAPQOVLNGVPSNFTYSPDTKVLDICVSLLMGQFLVSWC 1228
Db 841 ATAPQOVLNGVPSNFTYSPDTKVLDICVSLLMGQFLVSWC 883
RESULT 4
ID ADF47521
XX ADP47521 standard; protein; 950 AA.
AC ADF47521;
XX
XX
DT 26-FEB-2004 (first entry)
XX
XX Human GAA protein SEQ ID NO:24.
DE
DE
KW undersglycosylated targeted therapeutic; human; lysosome;
KW lysosomal targeting domain;
KW cation-independent mannose-6-phosphate receptor;
KW undersglycosylated therapeutic fusion protein; nephrotropic;
KW enzyme replacement therapy; cell therapy; gene therapy;
KW lysosomal storage disease; metabolic disease; enzyme defect;
KW Pompe disease; Tay-Sachs disease; Sandhoff disease; Fabry disease;
KW Gaucher disease; Krabbe disease; Wolman disease; Hurler syndrome;
KW Hunter syndrome; Sly syndrome; Schindler disease;
KW infantile sialic acid storage disease; Batten disease;
```

```
KW infantile neuronal ceroid lipofuscinosis; Ehlers-Danlos syndrome type VI;
KW congenital glycosylation disorder.
OS Homo sapiens.
PN WO2003102583-A1.
XX
PD 11-DEC-2003.
XX
XX 29-MAY-2003; 2003WO-US017211.
XX
XX 29-MAY-2002; 2002US-0384452P.
XX
XX 05-JUN-2002; 2002US-0386019P.
XX
XX 06-SEP-2002; 2002US-0408816P.
XX
XX 16-OCT-2002; 2002US-00272531.
XX
XX 06-FEB-2003; 2003US-0445734P.
XX
XX (SYMB-) SYMBIONTICS INC.
PA
XX
XX Lebowitz JH, Beverley SM, Sly WS;
XX
XX WPI; 2004-035473/03.
XX
XX N-PSDB; ADF47520.
XX
XX
XX Novel undersglycosylated targeted therapeutic comprising therapeutic agent
XX active in human lysosome, lysosomal targeting domain binds to Pompe
XX cation-independent mannose-6-phosphate receptor, for treating Pompe
XX disease.
XX
XX Example 13B; SEQ ID NO 24; 137pp; English.
```

The present invention describes an undersglycosylated targeted therapeutic (GT) which comprises: (a) a therapeutic agent that is therapeutically active in human lysosome; and (b) a lysosomal targeting domain that binds an extracellular domain of human cation-independent mannose-6-phosphate receptor (CM) and (i) does not bind a mutin in which amino acid 1572 of CM is changed from isoleucine to threonine; and (ii) binds the mutin with dissociation constant at least ten times the dissociation constant for binding CM. Also described: (1) an undersglycosylated therapeutic fusion protein (FP) comprising a therapeutic domain and a subcellular targeting domain that binds to an extracellular domain of a receptor on an exterior surface of a cell, and upon internalisation of the receptor, permits localisation of the therapeutic domain to a subcellular compartment where the therapeutic domain is therapeutically active; and (2) production of GT. GT has nephrotropic activity, and can be used in enzyme replacement therapy, cell therapy and gene therapy. The FP can be used for treating a lysosomal storage disease patient by administering FP to the patient. GT can be used for treating a patient by identifying a targeting moiety that binds CM in a mannose-6-phosphate independent manner, synthesising GT comprising a therapeutic agent that is therapeutically active in a mammalian lysosome and a targeting moiety that binds CM in a mannose-6-phosphate independent manner and administering GT to the patient, where the targeting moiety is identified by screening a nucleic acid or peptide library. GT is useful for treating metabolic disease, lysosomal storage diseases and associated enzyme defects such as Pompe disease, Tay-Sachs disease, Sandhoff disease, Fabry disease, Gaucher disease, Krabbe disease, Wolman disease, Hurler syndrome, Hunter syndrome, Sly syndrome, Schindler disease, infantile sialic acid storage disease, Batten disease, infantile neuronal ceroid lipofuscinosis, Ehlers-Danlos syndrome type VI and congenital disorders of glycosylation. The present sequence is used in the exemplification of the present invention.

Sequence 950 AA;

Query Match 72.4%; Score 4724.5; DB 8; Length 950;

Best Local Similarity 97.2%; Pred. No. 0; Mismatches 18; Indels 5; Gaps 1;

Matches 888; Conservative 3; Mismatches 18; Indels 5; Gaps 1;

QY 313 LEGRTELKELGYTVKHLQDLISGRISRAAEATGAHFGPRPRAVPTQCDVPPNSRFDCAPDK 372

Db 42 LSGSPVLEETHPAHQOQASRPGPR----DAQAHPGPRPRAVPTQCDVPPNSRFDCAPDK 96

```
QY 373 AITQCEARGCCYIPAKQGLQAGMGPWCFFPPSPSYKLENLSSEMGYTATLTRTT 432
Db 97 AITQCEARGCCYIPAKQGLQAGMGPWCFFPPSPSYKLENLSSEMGYTATLTRTT 156
QY 433 PTFPPKDLRLDVMETENRHLFTIKDPANRREYVPLETPRVHSRAPSPLYSVFSEE 492
Db 157 PTFPPKDLRLDVMETENRHLFTIKDPANRREYVPLETPRVHSRAPSPLYSVFSEE 216
QY 493 PFGVIVHRQLDGRVLLNTTVAFLFFADQFLQSLTSLPSQYITGLAEHLSPLMLSTWTRI 552
Db 217 PFGVIVHRQLDGRVLLNTTVAFLFFADQFLQSLTSLPSQYITGLAEHLSPLMLSTWTRI 276
QY 553 TLNWRDLAPTPGANLYGSHFYLALDGGSAHGVLFLNSNAMDVVLOPSPALSWRSTGGI 612
Db 277 TLNWRDLAPTPGANLYGSHFYLALDGGSAHGVLFLNSNAMDVVLOPSPALSWRSTGGI 336
QY 613 LDVYIFLGPEPKSVVQQYLDVGVGPFMPYVGLGFHLCRWGYSTALTROVVENMTRAHF 672
Db 337 LDVYIFLGPEPKSVVQQYLDVGVGPFMPYVGLGFHLCRWGYSTALTROVVENMTRAHF 396
QY 673 PLDVQWMDLDYMSRRDFTFNKDGFRDPPAMVQELHOGGRYMWIVDPAISSSGPAGSYR 732
Db 397 PLDVQWMDLDYMSRRDFTFNKDGFRDPPAMVQELHOGGRYMWIVDPAISSSGPAGSYR 456
QY 733 PYDEGLRGGVITNETQPLIGKWPBGSTAPPDFTNPTALAWEDMVAEFHDQVPDGLW 792
Db 457 PYDEGLRGGVITNETQPLIGKWPBGSTAPPDFTNPTALAWEDMVAEFHDQVPDGMW 516
QY 793 IDNNEPSNFRGSEDCPNNELENPPYVGVGGTLOAATICASSHQFLSTHYNLNLGY 852
Db 517 IDNNEPSNFRGSEDCPNNELENPPYVGVGGTLOAATICASSHQFLSTHYNLNLGY 576
QY 853 LTBALASHRALVKARGTRPFVISRSTFAGHGRVAGHTGVDVWSWQELASSVPEILQFNL 912
Db 577 LTBALASHRALVKARGTRPFVISRSTFAGHGRVAGHTGVDVWSWQELASSVPEILQFNL 636
QY 913 LGVPLVGADVCGFLGNTSEBELCVRWTLQGFYFPMRNHNSLLSLPQEPYFSFSPAQAAMR 972
Db 637 LGVPLVGADVCGFLGNTSEBELCVRWTLQGFYFPMRNHNSLLSLPQEPYFSFSPAQAAMR 696
QY 973 KALTLYALLPHLYTLFPHQAHVAGETVARPLFLEFPKDSSTWTVDHOLLWGEALLTPVL 1032
Db 697 KALTLYALLPHLYTLFPHQAHVAGETVARPLFLEFPKDSSTWTVDHOLLWGEALLTPVL 756
QY 1033 QAGKAEVTGYFPLGTWYDLQVPIEALGSLPPPPAPAPREPAIHSEGOWTLPLAPLDTINV 1092
Db 757 QAGKAEVTGYFPLGTWYDLQVPIEALGSLPPPPAPAPREPAIHSEGOWTLPLAPLDTINV 816
QY 1093 HLRAGYIIPQGPGLATTESRQQPMALAVALTGGEARGELFWDDGESLEVLERGAYTQV 1152
Db 817 HLRAGYIIPQGPGLATTESRQQPMALAVALTGGEARGELFWDDGESLEVLERGAYTQV 876
QY 1153 IFLARNNTIYNELVRVTSEGAGLQKVTVLGVATAPQOVLNGVPVSNFTYSPDTRKVL 1212
Db 877 IFLARNNTIYNELVRVTSEGAGLQKVTVLGVATAPQOVLNGVPVSNFTYSPDTRKVL 936
QY 1213 ICVSLLMGEQFLVS 1226
Db 937 ICVSLLMGEQFLVS 950
RESULT 5
ABM84315
ID ABM84315 standard; protein; 878 AA.
XX AC ABM84315;
XX DT 18-NOV-2004 (first entry)
XX DE Human diagnostic and therapeutic pprotein SEQ ID NO:4564.
XX KW gene therapy; human diagnostic and therapeutic polynucleotide; dithp.
XX
```

```
OS Homo sapiens.
PN WO2004023973-A2.
XX 25-MAR-2004.
XX 12-SEP-2003; 2003WO-US028227.
XX 12-SEP-2002; 2002US-0410259P.
PR 12-SEP-2002; 2002US-0410260P.
XX (INCY-) INCYTE CORP.
XX Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F,
PI Harthshorne TA, Suchorolski MT, Altus CM, Pitts SJ, Eider LV;
PI Mooney EM, Deleagane AM, Panesar IS, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstein EH;
PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vict UA, Kifton ES;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Patury S, Shi X, Suarez CJ;
XX WPI: 2004-329368/30.
DR N-PSDB; ACM42967.
XX
XX New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT in diagnosing a condition, disease or disorder associated with human
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT in gene mapping.
XX
XX Claim 27; Page; 190pp; English.
XX
XX The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human
CC polynucleotide and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorder, developmental disorders, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp protein of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX
XX Sequence 878 AA;
```

```
Query Match 65.7%; Score 4283.5; DB 8; Length 878;
Best Local Similarity 88.9%; Pred. No. 0;
Matches 814; Conservative 3; Mismatches 20; Indels 79; Gaps 2;
QY 313 LEGRTKELGYTVKKHLQDLSGRISARAEATGAHPRPRAVPTQCDVPNSRFDCAPOK 372
Db 42 LSGSSPVLSETHPAHQGASRPGPR-----DAQAHGPRPRAVPTQCDVPNSRFDCAPOK 96
QY 373 AITQCEARGCCYIPAKQGLQAGMGPWCFFPPSPSYKLENLSSEMGYTATLTRTT 432
Db 97 AITQCEARGCCYIPAKQGLQAGMGPWCFFPPSPSYKLENLSSEMGYTATLTRTT 156
QY 433 PTFPPKDLRLDVMETENRHLFTIKDPANRREYVPLETPRVHSRAPSPLYSVFSEE 492
Db 157 PTFPPKDLRLDVMETENRHLFTIKDPANRREYVPLETPRVHSRAPSPLYSVFSEE 216
QY 493 PFGVIVHRQLDGRVLLNTTVAFLFFADQFLQSLTSLPSQYITGLAEHLSPLMLSTWTRI 552
Db 217 PFGVIVHRQLDGRVLLNTTVAFLFFADQFLQSLTSLPSQYITGLAEHLSPLMLSTWTRI 276
QY 553 TLNWRDLAPTPGANLYGSHFYLALDGGSAHGVLFLNSNAMDVVLOPSPALSWRSTGGI 612
|||||
```

Db 277 TLWNRDLAPTGANLYGSHPPFYALBEDGSGAHGVFLNLSNANDVVLQSPALSWRSTGGI 336
QY 613 LDVYIFLGPPEKSVVQOYLDDVGVGPPMPYWGILGFLHCRWGSSTAITRQVVENMTRAHF 672
Db 337 LDVYIFLGPPEKSVVQOYLDDVGVGPPMPYWGILGFLHCRWGSSTAITRQVVENMTRAHF 396
QY 673 PLDVQNDLDYMDSRDRDFTFNKDGFRDFFAMVQELHGGGRYMMIVDPAISSGSPAGSYR 732
Db 397 PLDVQNDLDYMDSRDRDFTFNKDGFRDFFAMVQELHGGGRYMMIVDPAISSGSPAGSYR 456
QY 733 PYDEGLRRGVFTNETGQPLIGKWPGSTAFDPFTNPTALAWEDMVAEFHDOVDPDGLW 792
Db 457 PYDEGLRRGVFTNETGQPLIGKWPGSTAFDPFTNPTALAWEDMVAEFHDOVDPDGLW 516
QY 793 IDNNEPSNIRGSDCPCNNLENPYPVGVGGTILQAATICASSHQFLSTHYNLHLYG 852
Db 517 IDNNEPSNIRGSDCPCNNLENPYPVGVGGTILQAATICASSHQFLSTHYNLHLYG 576
QY 853 LTEAIAASHRALVKARTRPFVISRSTFAGHRYAGHWTGDVWSSWQLASSVPEILQFNL 912
Db 577 LTEAIAASHRALVKARTRPFVISRSTFAGHRYAGHWTGDVWSSWQLASSVPEILQFNL 636
QY 913 LGVPLVGADVCGFLGNTSEELCVRWTLQGFAPFPMENHNSLSLPOEPYFSEPAQQAAR 972
Db 637 LGVPLVGADVCGFLGNTSEELCVRWTLQGFAPFPMENHNSLSLPOEPYFSEPAQQAAR 696
QY 973 KALTLYALLPHLYTLFHOAHVAGETVARPLFLEFPKDSSTWTVDHQLLWGEALLITPVL 1032
Db 697 KALTLYALLPHLYTLFHOAHVAGETVARPLFLEFPKDSSTW----- 738
QY 1033 QAGKAEVTVFPLGTWYDILQTVPIEALGSLPPPPAAPREPAIHSEQWTLPAPLDTINV 1092
Db 739 -----TINV 742
QY 1093 HLRAGYIIPLOGPLTTTSROOPMALAVALTKGGEARGELFWDDGESLEVLERGAYTOV 1152
Db 743 HLRAGYIIPLOGPLTTTSROOPMALAVALTKGGEARGELFWDDGESLEVLERGAYTOV 802
QY 1153 IFLARNNTIVNELVRVTSAGAGLQKQVTLGVATAPQOVLNGVPSNFTYSPDTKVLD 1212
Db 803 IFLARNNTIVNELVRVTSAGAGLQKQVTLGVATAPQOVLNGVPSNFTYSPDTKVLD 862
QY 1213 ICVSLLMGQFLVSWC 1228
Db 863 ICVSLLMGQFLVSWC 878

RESULT 6
ABM84314
ID ABM84314 standard; protein; 904 AA.
XX
AC ABM84314;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human diagnostic and therapeutic pprotein SEQ ID NO:4563.
XX
KW gene therapy; human diagnostic and therapeutic polynucleotide; dithp.
XX
OS Homo sapiens.
XX
PN WO2004023973-A2.
XX
PD 25-MAR-2004.
XX
PP 12-SEP-2003; 2003WO-US028227.
XX
PR 12-SEP-2002; 2002US-0410259P.
PR 12-SEP-2002; 2002US-0410260P.
XX
PA (INCY-) INCYTE CORP.
XX
PI Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;

PI Harthshorne TA, Suchorolski MT, Alcus CM, Pitts SJ, Elder LV;
PI Mooney EM, Deleane AM, Panesar IS, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;
PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Patry S, Shi X, Suarez CU;
XX
XX WPI; 2004-329368/30.
DR N-PSDB; ACN42966.
XX
XX New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT in diagnosing a condition, disease or disorder associated with human
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT in gene mapping.
XX
PS Claim 27; Page; 190pp; English.
XX
XX The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human
CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorder, developmental disorder, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp protein of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX
SQ Sequence 904 AA;

Query Match 65.3%; Score 4260.5; DB 8; Length 904;
Best Local Similarity 86.4%; Pred. No. 0;
Matches 814; Conservative 3; Mismatches 20; Indels 105; Gaps 3;

QY 313 LEGRTKELGYTKVKKHQLDLSGRISRAAEATGAHGRPRAPVPTCCDVPNRRFDCAPDK 372
Db 42 LSGSPVLEETHPAHQOGASRPGPR-----DAQAHGPRAPVPTCCDVPNRRFDCAPDK 96
QY 373 AITQCEARGCCYIPAKQGLQGAQWQPWCFPPSPYSYKLENLSSESMGYTALTRTT 432
Db 97 AITQCEARGCCYIPAKQGLQGAQWQPWCFPPSPYSYKLENLSSESMGYTALTRTT 156
QY 433 PTFPPKDIILTLRLDVMETENRLHFTIKDPANRRYEVPLETPRHVSRAAPSPLYSEFSEE 492
Db 157 PTFPPKDIILTLRLDVMETENRLHFTIKDPANRRYEVPLETPRHVSRAAPSPLYSEFSEE 216
QY 493 PEGVIVRQLDGRVLLNTTVAFLFADQFLQSLPSQYITGLAEHLSPLMLSTSWTRI 552
Db 217 PEGVIVRQLDGRVLLNTTVAFLFADQFLQSLPSQYITGLAEHLSPLMLSTSWTRI 276
QY 553 TLWNRDLAPTGANLYGSHPPFYALBEDGSGAHGVFLNLSNANDVVLQSPALSWRSTGGI 612
Db 277 TLWNRDLAPTGANLYGSHPPFYALBEDGSGAHGVFLNLSNANDVVLQSPALSWRSTGGI 336
QY 613 LDVYIFLGPPEKSVVQOYLDDV-----GYFMPPPYWGILG 546
Db 337 LDVYIFLGPPEKSVVQOYLDDVVGACSLAAAPAPRLPPPSLMKSLALACGYFMPPPYWGILG 396
QY 647 FHLCEWGYSSTAITRQVVENMTRAHFPLDVQNDLDYMDSRDRDFTFNKDGFRDFAWVOE 706
Db 397 FHLCEWGYSSTAITRQVVENMTRAHFPLDVQNDLDYMDSRDRDFTFNKDGFRDFAWVOE 456
QY 707 LHQGGRRYMMIVDPAISSGSPAGSYRYPYDEGLRRGVFTINETGQPLIGKWPGSTAFPDF 766
Db 457 LHQGGRRYMMIVDPAISSGSPAGSYRYPYDEGLRRGVFTINETGQPLIGKWPGSTAFPDF 516

QY 767 TNPATLAWMEDVVAEFDHQPDPGLWDMNPSNFIRGSEDCGCPNNELENPPYVPGVGG 826
 DB 517 TNPATLAWMEDVVAEFDHQPDPGLWDMNPSNFIRGSEDCGCPNNELENPPYVPGVGG 576
 QY 827 TLQAATICASSHQFLSTHYNLNLXGLTEAASHRALVKARGTRPPIVSRSTFAGHGRYA 886
 DB 577 TLQAATICASSHQFLSTHYNLNLXGLTEAASHRALVKARGTRPPIVSRSTFAGHGRYA 636
 QY 887 GHWTGDVWSSWEQLASSVPEILQFNLLGVPLVGADVCGFLGNTSEELCVRWTLQAGAFYPP 946
 DB 637 GHWTGDVWSSWEQLASSVPEILQFNLLGVPLVGADVCGFLGNTSEELCVRWTLQAGAFYPP 696
 QY 947 MRNHNLSLLSQBPYPYSEPAQAMRKALTRYALLPHLYTLFHOAHVAGETVARPLFLLE 1006
 DB 697 MRNHNLSLLSQBPYPYSEPAQAMRKALTRYALLPHLYTLFHOAHVAGETVARPLFLLE 756
 QY 1007 FPKDSSTMTVDHQLLWGEALLTPVLQAGKAEVTGYFPLGTWYDLQTVFIEALGSLPPPP 1066
 DB 757 FPKDSSTW----- 764
 QY 1067 AAPREPAIHSEGOWTLPAPLDTINVHLRAGYIIPQGPGLTTTSSRQOPMALAVALTKG 1126
 DB 765 -----TINVHLRAGYIIPQGPGLTTTSSRQOPMALAVALTKG 802
 QY 1127 GEARGELFWDDGESLEVLERGAYTOVIFLARNTIIVNELVRVTSEGAGLOLOKVTVLQVA 1186
 DB 803 GEARGELFWDDGESLEVLERGAYTOVIFLARNTIIVNELVRVTSEGAGLOLOKVTVLQVA 862
 QY 1187 TAPQOVLNGVPSNFTYSPDRTKVLDCVSLLMGEQFLYSWC 1228
 DB 863 TAPQOVLNGVPSNFTYSPDRTKVLDCVSLLMGEQFLYSWC 904

RESULT 7

ABB57174
 ID ABB57174 standard; protein; 953 AA.
 XX
 AC ABB57174;
 XX
 DT 07-MAR-2002 (first entry)
 XX
 DE Mouse ischaemic condition related protein sequence SEQ ID NO:422.
 XX
 KW Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
 KW vasoepaetic ischaemia; ischaemic condition; ischaemic disease.
 XX
 OS Mus musculus.
 XX
 PN WO200188188-A2.
 XX
 PD 22-NOV-2001.
 XX
 PF 18-MAY-2001; 2001WO-JP004192.
 XX
 PR 18-MAY-2000; 2000JP-00145977.
 XX
 PA (UYN1-) UNIV NIHON SCHOOL JURIDICAL PERSON.
 XX
 PI Ishikawa K, Asai S, Takahashi Y, Negata T, Ishii Y;
 XX
 DR WPI; 2002-034733/04.
 DR N-PSDB; ABI99468.
 XX
 PT Examining the ischemic condition (e.g. occlusive ischemia) by measuring
 PT expression levels of particular genes defined in the specification or by
 PT determining the expression profile of a gene group comprising these
 PT genes.
 XX
 PS Claim 2; Page 1154-1158; 2690pp; English.
 XX
 CC The present invention describes a method for examining ischaemic
 CC conditions, comprising measuring the expression levels of particular
 CC genes (i) in a test sample or determining the expression profile of a

CC gene group in the sample comprising genes selected from (1). The method
 CC is useful for examining the ischaemic condition (e.g. compressive
 CC ischaemia, occlusive ischaemia or vasoepaetic ischaemia) by measuring
 CC expression levels of particular genes (ABI99202 to ABI99912, encoding the
 CC protein sequences in ABB57020 to ABB57374) or by determining the
 CC expression profile of a gene group comprising these genes. The expression
 CC levels or expression profiles produced by these genes are used as an
 CC indicator when screening for ischaemic condition-improving drugs or
 CC therapeutics for ischaemic diseases. ABI99913 and ABI99914 represent PCR
 CC primers for a mouse ischaemic condition related sequence, which are used
 CC in the exemplification of the present invention
 XX

SQ Sequence 953 AA;

Query Match 60.9%; Score 3974; DB 5; Length 953;
 Best Local Similarity 79.8%; Pred. No. 0;
 Matches 739; Conservative 73; Mismatches 96; Indels 18; Gaps 4;
 QY 320 LGYTVKKHL---QDL-----SGRISRAAEATGAHPG-----RPRVPTQCDVPP 362
 DB 27 LGHLMRLMLLPQDLHSSSLGKWTYRPHHOGYEPGFLHTQEQTEQKPEAPTQCDVPP 86
 QY 363 NSRFDCAPDKAITQCEARGCCYIPAKQGLQAGQWQPFCFPFPSPYPSYKLENLSSEM 422
 DB 87 NSRFDCAPDKGLSQCEARGCCYIPAGQVLKEPQIGQPCWCFPPSPYPSYRLENLSSTES 146
 QY 423 GYTATLTRTPTFFPKDILTLRLDVMETENLHFTIKDPANRREYEVPLETPRVHGRAPS 482
 DB 147 GYTATLTRTPTFFPKDVLTLQLEVLMTEDSLHFKIKDPASKRYEVPLETPRVLSQAPS 206
 QY 483 PLYSVFSEEPFGVIVHQLDGRVLLNTTVAFLPADQLQLSTLPSQYITGLAEHLSP 542
 DB 207 PLYSVFSEEPFGVIVRKLGRVLLNTTVAFLPADQLQLSTLPSQAHIITGLGHLSP 266
 QY 543 LMLSTSWTRITLWNRDLAPTGCANLYGSHPPFLALEDCGSAHGVPFLNANMDVVLQPPS 602
 DB 267 LMLSTDMARITLWNRDTPPSQGTNLYGSHPPFLALEDCGLAHGVFLNANMDVVLQPPS 326
 QY 603 ALSMRSTGGILDVYIFLQPEPKSVVQQLDVVGYPFMPYPWGLGHLCRWGSYSTAIRQ 662
 DB 327 ALTWSTGGILDVYVFLGPEPKSVVQQLDVVGYPFMPYPWGLGHLCRWGSYSTAIRQ 386
 QY 663 VVENNTRAHFPLDVQWNLDDYMSRRDPTFNKDGFRDPPAMQVQELHOGGRRYMMIVDPAI 722
 DB 387 VVENNTRTHFPLDVQWNLDDYMDARDEFTFNQDSFADFPDMVRDVHQQGRRYMMIVDPAI 446
 QY 723 SSSGPAGSVRPYDEGLRGVFTTNETGOLICKVMPGSTAFDPDFTNPTALAWMEDVVAEF 782
 DB 447 SSAGPAGSVRPYDEGLRGVFTTNETGOLICKVMPGSTAFDPDFTNPTALAWMEDVVAEF 506
 QY 783 HDQVPFDGLWIDMNEPSNPIRGSEDCGCPNNELENPPYVPGVGGTLQAAATICASSHQFLS 842
 DB 507 HQQVPFDGWLMDMNEPSNPIRGSEDCGCPNNELENPPYVPGVGGTLQAAATICASSHQFLS 566
 QY 843 THYNLNLXGLTEAASHRALVKARGTRPPIVSRSTFAGHGRYAGHWTGDVWSSWQLAS 902
 DB 567 THYNLNLXGLTEAASHRALVKARGTRPPIVSRSTFAGHGRYAGHWTGDVWSSWQLAS 636
 QY 903 SVPEILOENLLGVPLVGADVCGFLGNTSEELCVRWTLQAGAFYPMRNNHNSLLSLPOEPS 962
 DB 627 SVPEILOENLLGVPLVGADVCGFLGNTSEELCVRWTLQAGAFYPMRNNHNSLLSLPOEPS 686
 QY 963 FSEPAQAMRKALTRYALLPHLYTLFHOAHVAGETVARPLFLFEPKDSSTMTVDHQLLW 1022
 DB 687 FSEPAQAMRKALTRYALLPHLYTLFHOAHVAGETVARPLFLFEPKDSSTMTVDHQLLW 746
 QY 1023 GBALLITPVLOAGKAEVTGYFPLGTWYDLQTVFIEALGSLPPPPAAR-EPAIHSEGGQV 1081
 DB 747 GPALLITPVLEPGKTEVTGYFPKGTWYNNQVSVDSLGLTLPSPSSASSFRSAVQSGQWL 806
 QY 1082 TLPAELDTINVHLRAGYIIPQGPGLTTTSSRQOPMALAVALTKGEARGELFWDDGESL 1141
 DB 807 TLEAPLDTINVHLRAGYIIPQGPGLTTTSSRQOPMALAVALTKGEARGELFWDDGESL 866

QY 1142 EVLERGAYTQVIFLARNNTIVNELVRVTSEGAGLQKQVTLGVATAPQOVLNSGVPSN 1201
DB 867 AVLEHGYTLVFSAKNTIVNKLVRVTEGAELQLKEVTLGVATAPQVLSNGIPVSN 926
QY 1202 FTYSPDTKVLDICVSLLMGEQFLVSW 1227
DB 927 FTYSPDNKSLAIPVSLLMGELFQISW 952

RESULT 8

ABM84316
ID ABM84316 standard; protein; 788 AA.

AC ABM84316;

XX 18-NOV-2004 (first entry)

DE Human diagnostic and therapeutic pprotein SEQ ID NO:4565.

XX gene therapy; human diagnostic and therapeutic polynucleotide; dithp.

XX Homo sapiens.

OS WO2004023973-A2.

PN 25-MAR-2004.

XX 12-SEP-2003; 2003WO-US028227.

XX 12-SEP-2002; 2002US-0410259P.

PR 12-SEP-2002; 2002US-0410260P.

XX (INCY-) INCYTE CORP.

XX Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F, Hartshorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV, Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP, Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Geratin EH, Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL, Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES, Xu Y, Kwong M, Pollicky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D, Fatury S, Shi X, Suarez CJ;

XX WPI: 2004-329368/30.

DR N-PSDB; ACM42968.

XX New diagnostic and therapeutic polynucleotides and polypeptides, useful in diagnosing a condition, disease or disorder associated with human molecules, e.g. autoimmune or inflammatory disorders, -in gene therapy or in gene mapping.

XX Claim 27; Page; 190pp; English.

XX The invention relates to novel diagnostic and therapeutic polynucleotides selected from one of the 2722 sequences defined in the specification. A polynucleotide of the invention may have a use in gene therapy. The human diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be used to diagnose a particular condition, disease or disorder associated with human molecules, e.g. cell proliferative disorders, autoimmune/inflammatory disorder, developmental disorder, endocrine disorder, neurological disorders, gastrointestinal disorders, or infections caused by virus, bacteria, fungi or parasite. The dithp molecules may also be used in genetic mapping, in identifying individuals from minute biological samples, in detecting single nucleotide polymorphisms, as molecular weight markers, and for somatic or germline gene therapy. The present sequence represents a dithp protein of the invention. Note: The sequence data for this patent is not represented in the printed specification, but was obtained in electronic format directly from WIPO at www.wipo.int/pct/en/sequences/listing.htm

XX Sequence 788 AA;

ADJ83134 standard; protein; 707 AA.

XX

Query Match 57.7%; Score 3763.5; DB 8; Length 788;
Best Local Similarity 79.0%; Pred. No. 1e-310;
Matches 724; Conservative 3; Mismatches 20; Indels 169; Gaps 2;

QY 313 LEGRTKEIGYVKKHLQDLSGRISARABAEATGAHGPRAVPTQCDVPPNSRFFCAPDK 372
DB 42 LSGSSPVLSEETHPAHQOGASRPGR-----DAQAHGPRPRAVPTQCDVPPNSRFFCAPDK 96
QY 373 AITQEQCEARGCCYIPAKQGLQGAQGWPCFFPPSPYSYKLENLSSEMGTATLTRIT 432
DB 97 AITQEQCEARGCCYIPAKQGLQGAQGWPCFFPPSPYSYKLENLSSEMGTATLTRIT 156
QY 433 PTFPPKIDILRLDVMETENRLHFTIKDPANRRYEVPLETPRVHSRAPSPLYSVEFSE 492
DB 157 PTFPPKIDILRLDVMETENRLHFTIKDPANRRYEVPLETPRVHSRAPSPLYSVEFSE 216
QY 493 PFGVIVRQLDGRVLLNTTVAFLPFADQFLQSLPSQYITGLAEHLSPLMLSTSWTRI 552
DB 217 PFGVIVRQLDGRVLLNTTVAFLPFADQFLQSLPSQYITGLAEHLSPLMLSTSWTRI 276
QY 553 TLWNRDLAPTGANLYGSHPPFYLALEDGSAHGVELLNSNAMDVVLQSPALSWSTGGI 612
DB 277 TLWNRDLAPTGANLYGSHPPFYLALEDGSAHGVELLNSNAMDVVLQSPALSWSTGGI 336
QY 613 LDVYIFLGPPEKSVVQQYLDVVGYFMPFYMGLGFHLCRWGYSSTAITRQVVENNTRAHP 672
DB 337 LDVYIFLGPPEKSVVQQYLDVVGYFMPFYMGLGFHLCRWGYSSTAITRQVVENNTRAHP 396
QY 673 PLDVQWNLDYMSRRDFTFNKDGPRDFPMVQELHQGRRYMMIVDPAISSGGPAGSYR 732
DB 397 PLDVQWNLDYMSRRDFTFNKDGPRDFPMVQELHQGRRYMMIVDPAISSGGPAGSYR 456
QY 733 PYDEGLRRGVFITNETGQPLIGKWPGSTAPDFNTALTAWMEDVAFHDFQVDFDGLW 792
DB 457 PYDEGLRRGVFITNETGQPLIGKWPGSTAPDFNTALTAWMEDVAFHDFQVDFDGLW 516
QY 793 IDMEPSNFIRSGEDGCNNELNPPYVGVGGTLOAATICASSHQFLSTHYNLHLYG 852
DB 517 IDMEPSNFIRSGEDGCNNELNPPYVGVGGTLOAATICASSHQFLSTHYNLHLYG 576
QY 853 LTEATASHRALVKARGTTPFVISRSTFAGHGYAGHWTGDVWSSWEOLASSVPEILOLNL 912
DB 577 LTEATASHRALVKARGTTPFVISRSTFAGHGYAGHWTGDVWSSWEOLASSVPEILOLNL 636
QY 913 LCVPLVGADVCGFLGNTSEELCVMTQLGAFYPMFRNHNLSLSPQEPYSFSEPAQAMR 972
DB 637 LCVPLVGADVCGFLGNTSEELCVMTQLGAFYPMFRNHNLSLSPQEPYSFSEPAQAMR 696
QY 973 KALTURYALLPHLYTLFHOAHVAGETVARPLFEPFKDSSTWTVDHQLLMGALLITPVL 1032
DB 697 KALT----- 700
QY 1033 QAKKAEVTCYPLGTWYDLOTVPIEALGSLPPPPAPPAHPAIHSEGWVTLPAPLDTINV 1092
DB 701----- 700
QY 1093 HLRAGVIIPLOGPGLTTSRQPPMALAVALTKGGEARGELFDDGSELEVLERGAYTQV 1152
DB 701-----LEVLERGAYTQV 712
QY 1153 IFLARNNTIVNELVRVTSEGAGLQKQVTLGVATAPQOVLNSGVPSNFTYSPDTKVLD 1212
DB 713 IFLARNNTIVNELVRVTSEGAGLQKQVTLGVATAPQOVLNSGVPSNFTYSPDTKVLD 772
QY 1213 ICVSLLMGEQFLVSWC 1228
DB 773 ICVSLLMGEQFLVSWC 788

RESULT 9

ADJ83134

ID ADJ83134 standard; protein; 707 AA.

XX

536 FPDAAETDIDRQFLWGSALLNAPVLEPGATSKAYLPGRWYDLYTGAGEA----- 587
1067 AAPREPAIHSEGOWTLAPLDITINHLRAGYIIPLOQPLATTESRQOPMALAVATKG 1126
588 -----SRGNVTLSPDKIPVHVRRGSIIPQEPALTTESRDNPFFHLLVALDDN 638
1127 GEARGELFWDDGESLEVLERGAYQVIFLARNTTIVNELRVVT---SEGAGLQKQVTVL 1183
639 GTASGELYLDDGESIOT-QRGDLYLLVQFSANNNTLGTGV-VTGYYKNSNTLTLEKITL 696
1184 GVATAPOQV 1192
697 GVGNEPAAV 705

RESULT 10
ABM84317
ID ABM84317 standard; protein; 638 AA.
XX ABM84317;
AC ABM84317;
DT 18-NOV-2004 (first entry)
XX Human diagnostic and therapeutic pprotein SEQ ID NO:4566.
DE Human diagnostic and therapeutic pprotein SEQ ID NO:4566.
XX gene therapy; human diagnostic and therapeutic polynucleotide; dithp.
XX Homo sapiens.
XX OS Homo sapiens.
XX WO2004023973-A2.
XX 25-MAR-2004.
XX 12-SEP-2003; 2003WO-US028227.
XX 12-SEP-2002; 2002US-0410259P.
XX 12-SEP-2002; 2002US-0410260P.
XX (INCY-) INCYTE CORP.
XX Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F,
PI Harthorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;
PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Paneser SR, Wang X, Au AP, Gerstin EH;
PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES;
PI Xu Y, Kwong M, Pollocky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Patury S, Shi X, Suarez CJ;
XX WPI; 2004-329368/30.
XX N-PSDB; ACN42969.
XX New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT in diagnosing a condition, disease or disorder associated with human
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT in gene mapping.
XX Claim 27; Page; 190pp; English.
XX The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human
CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorder, developmental disorder, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp protein of the
CC invention. Note: The sequence data for this patent is not represented in

CC the printed specification, but was obtained in electronic format directly
CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX
SQ Sequence 638 AA;
Query Match 30.8%; Score 2007.5; DB 8; Length 638;
Best Local Similarity 49.7%; Pred. No. 3.2e-161;
Matches 465; Conservative 19; Mismatches 94; Indels 357; Gaps 22;
313 LEGRTKELGYTVKHLQDLSGRISARAEATGAHPRPRAVPTCCDVPNRSRFDCAPOK 372
42 LSGSPVLEETHPAHQOGASRGP-----DAQAHGPRPRAVPTCCDVPNRSRFDCAPOK 96
373 AITQEQCEARGCCYIPAKQLOGQAGQWPCFPFPPSPYKLENLSSEMGMYTALTRTT 432
97 AITQEQCEARGCCYIPAKQLOGQAGQWPCFPFPPSPYKLENLSSEMGMYTALTRTT 156
433 PTFPPKDIILRLDVMETENRLHFTIKDPANRRYEVPLETPRVHSRAPSPLYSVEFSEE 492
157 PTFPPKDIILRLDVMETENRLHFTIKDPANRRYEVPLETPRVHSRAPSPLYSVEFSEE 216
493 PFGVIVRQDGRVLLNTTVAPLFADQFLOLSTLPSQYITGLAEHLSPMLSTSWTRI 552
217 PFGVIVRQDGRVLLNTTVAPLFADQFLOLSTLPSQYITGLAEHLSPMLSTSWTRI 276
553 TLWNRDLAPTCANLYGSHPPFYLALEDGSGAHGVFLNLSNMDVVLQSPALSRWSTGGI 612
277 TLWNRDLAPTCANLYGSHPPFYLALEDGSGAHGVFLNLSNMDVVLQSPALSRWSTGGI 336
613 LDVYIFLGPPEKSVVQYLDVVVG-----YPMPPYVGLGFLCRWGYSTAITROV 663
337 LDVYIFLGPPEKSVVQYLDVVVGACSLAAAPAPRLPP-----PSL 377
664 VENMTFAHFPDLVDQVNDLYDMSRDTFTNKGDFRFPAMVQELHOGGRRYMMIVDPAIS 723
378 MKSALACRIPV-----SYRPPYDEGLRGVFTNETGOLIGKVMFGSTAFPDTNPTALAWMEDV 779
724 SSGPAG-----SYRPPYDEGLRGVFTNETGOLIGKVMFGSTAFPDTNPTALAWMEDV 416
399 PV-PLGLLLRHYP-----PGGGE----- 451
780 AEFHQDPFDGLWIDMNEFSPFIRSEGCNNELENPPYVGVVGGTTLQAATICASSHQ 839
417 ---HDQGPLP-----PGRPVERP----- 431
840 FLSTHYNLHNLGLTEAISHRALVKARGTRFVLSRSTFAGHGRYAGHWTGDNVSSWEQ 899
432 -----GLHGLPEGL-----HVQQGW----- 446
900 LASSVPEILQFNLLGVPLVGADVCGFLGNTSELCVRWTLQGFAPFPMNHNLSLSLPOE 959
447 -----LPGLP----- 451
960 PVSFSEPAQAAMRKALTURLYALLPHLYTLFHOAHVAGETVARPLFLEPKDSSTWVDHQ 1019
452 -----GHGAGAAPGRPA-LHDDRGSC-----HQ 473
1020 LLWGEALLITPVLOAKAEVTGYF-----PLGTWVDLQTVPIEALGSLPPPPAAPREPA 1073
474 QL-GPCREL-QALRRGSAB-----GGFHHQDRPADMEGWARV-----HCLPRLHQ 518
1074 IHSQGQWVTLPAPLDTINVHLRAGYIIPLOQPLATTESRQOPMALAVATKGGEARGEL 1133
519 PHS-----PGL-----QPMALAVATKGGEARGEL 543
1134 FWDDESLEVLERGAYQVIFLARNTTIVNELRVTSRAGLQLOKQVTVLGVATAPQOVL 1193
544 FWDDESLEVLERGAYQVIFLARNTTIVNELRVTSRAGLQLOKQVTVLGVATAPQOVL 603
1194 SNGVPSNFTYSPDKVLDICVSLAMGQFLVSWC 1228
604 SNGVPSNFTYSPDKVLDICVSLAMGQFLVSWC 638

RESULT 11
AD019560
ID ADO19560 standard; protein; 1857 AA.
XX AC ADO19560;
XX DT 12-AUG-2004 (first entry)
XX DE Human PRO polypeptide #245.
XX KW Human; PRO; immune related disorder; systemic lupus erythematosus;
KW rheumatoid arthritis; osteoarthritis; juvenile chronic arthritis;
KW systemic sclerosis; Sjogren's syndrome; vasculitis; sarcoidosis;
KW autoimmune haemolytic anaemia; autoimmune thrombocytopenia; thyroiditis;
KW diabetes mellitus; renal disease; demyelinating disease;
KW central nervous system; peripheral nervous system;
KW demyelinating polyneuropathy; Guillain-Barre syndrome;
KW chronic inflammatory demyelinating polyneuropathy.
XX OS Homo sapiens.
XX PN WO2004043361-A2.
XX PD 27-MAY-2004.
XX PF 06-NOV-2003; 2003WO-US035268.
XX PR 08-NOV-2002; 2002US-0425235P.
XX PA (GETH) GENENTECH INC.
XX PI Fong S, Dennis K, Clark H, Chiu H, Schoenfeld J, Williams PM;
PI Wood WI, Wu TD;
XX DR WPI; 2004-420067/39.
XX DR N-PSDB; AD019559.
XX PT Novel PRO polypeptide e.g., PRO69614, PRO71106, or PRO86388 useful for
PT treating an immune related disorder such as systemic lupus erythematosus,
PT rheumatoid arthritis, osteoarthritis, juvenile chronic arthritis or
PT spondyloarthritis.
XX PS Claim 7; SEQ ID NO 490; 1731pp; English.
XX CC The invention relates to human PRO polypeptides and the polynucleotides
CC encoding them. The polypeptides and polynucleotides are useful for
CC treating and diagnosing immune related disorders in mammals. The immune
CC related disorders include systemic lupus erythematosus, rheumatoid
CC arthritis, osteoarthritis, juvenile chronic arthritis, systemic
CC sclerosis, Sjogren's syndrome, vasculitis, sarcoidosis, autoimmune
CC haemolytic anaemia, autoimmune thrombocytopenia, thyroiditis, diabetes
CC mellitus, immune-mediated renal disease, demyelinating diseases of the
CC central or peripheral nervous system, demyelinating polyneuropathy,
CC Guillain-Barre syndrome and chronic inflammatory demyelinating
CC polyneuropathy. This sequence represents a human PRO polypeptide of the
CC invention.
XX SC Sequence 1857 AA;
Query Match 29.1%; Score 1900.5; DB 8; Length 1857;
Best Local Similarity 43.7%; Pred. NO. 2.7e-151;
Matches 394; Conservative 133; Mismatches 328; Indels 47; Gaps 21;
QY 348 RGRPRVPTQCDVP---PNSRDCAPDKAITQCEARGCCYIPAKQGLQAGQMPWCF 404
DB 78 PGTGTTPVSACPVVNERINCIPDPPTKATCDQRCGWNP-----QGA-VSPWCY 131
QY 405 FPPSPSYKLE-NLSSSEMGYATLTRTPT--FPKDLTLRLDVMETENRLHPTIKD 461
DB 132 YKXNH-SYHVEGNLVNAGTARL-KNLPSPVFGSNVDNVLITAYQTSNRHFKLTD 189
QY 462 PANRRYVELETPRVHS--RAPSPLYSVFSEEPFGVIVHRQDGRVLLNTTTPVLPFFAD 519

Db 190 QTNRFEPHEHVQSFSGNAAASLTYYQVEISQPFISIKVTRRSNNRVLFDSSIGPLLPAD 249
QY 520 QFLQLSTSLPSQYITGLAEHL-SPLMLSTSWTRIILMNRDLAPT-PGANLYGSHPFYAL 577
Db 250 QFLQLSTSLPSQYITGLAEHL-SPLMLSTSWTRIILMNRDLAPT-PGANLYGSHPFYAL 577
QY 578 ED-GGSAHGAVFLLNSAMDMVLQSPALSWRTGGILDVYIFLGPEPKSVVQOYLVDVGY 636
Db 310 EDASGLSGFVFLMNSNAMEVVLQAPAPAITRTIGGILDFYVFLGNTPEQVQVYLELIGR 369
QY 637 PMPYPVGLGPHLCBWSYSSTAITQVVENMTTRAHFPDLDVQNDLDYMDRDEFTFNKDG 696
Db 370 PALPSYALGFLHLSYEGTLDNMEVVERNAALQPYDVQHADIDYMDRDEFTYDSDV 429
QY 697 PRDFFAMVQELHGGRRYRMIIVDPAIS--SSGPAGSYRYPYDEGLRRGVFTTNTG-QPLIG 754
Db 430 FKGFPEFVNEHLNNGOKLVIIVDPAISNNSSSKYPGYDRGDMKIWNSSDGVTPFLIG 489
QY 755 KVMGSTAPPDFTNPTALAWEDMVAEHDQVPPDGLWIDMNEPSPNIFRGSDEGCGCNEL 814
Db 490 EVMPGQTVFPDYNPCAVMTKEPELPHNQVEFDGIWIDMNEVSNFVDSVSGCSTNNL 549
QY 815 ENPPYVPGVGTLOAATICASSHQFLSTHNLNLYGLTEALASHRALVKA-RGRPFV 873
Db 550 NNPPFTPRILDGFLCKLCMDAVQHWGQXYDIHNLGYSMATAAATAKTVPPNKRSTI 609
QY 874 ISRSTFAGHGRVAGHWTGDVSSWQSLASSPEIILQFNLLGVPLVGLADYCGFLGNTSEEL 933
Db 610 LTRSTFAGSGKFAAHLWGDNTATWDDLRSIPGVLEFNLFGIPMVGPDICGFALDTPPEL 669
QY 934 CVRTQLGAFYPMNRNHSLLSLPQEPYSFSPA--QQAMRKALTRYALLPHLYTLFHQ 991
Db 670 CRRWQLGAFYPMNRNHSLLSLPQEPYSFSPA--QQAMRKALTRYALLPHLYTLFHQ 991
QY 992 AHVAGETVARPLFLFPKDSSTWTDHQLLWGEALLITPVLAGKAEVTGYEPLGTWYDL 1051
Db 730 AHSRGDTVARPLFLFPKDSSTWTDHQLLWGEALLITPVLAGKAEVTGYEPLGTWYDL 1051
QY 1052 QTVPIEALQSLPPPAAPREPAIHSEGGWVTLPAPLDITNVHLRAGYIIPLOQPGLTITTE 1111
Db 790 ET-----GS-----QVRWRKQKVENELPDKLGLHLRGYIFPTQOPNTTILA 832
QY 1112 SRQPMALAVALTKEGARGELFWDGSEVLERGAYTQVIF-LARNNTIIVNELVRVTS 1170
Db 833 SRKNLGLIILALDENKEAGELFWDGGETKDTVANKVYLCEFSVTQNRLEVNISQSTYK 892
QY 1171 EGAGLQLOKVVVLGVATAPQ-OVLSNGVPS---NFTTSPDTKVLDIC-VSLLMGEQFLV 1225
Db 893 DPNNLAFNEIKILGTEEPSNVTGVNGVPSQTSPTVTYDNLKVAITIDILLGSEAYTV 952
QY 1226 SW 1227
Db 953 EW 954
RESULT 12
AAW74090
ID AAW74090 standard; protein; 1827 AA.
XX AC AAW74090;
XX DT 04-MAY-1999 (first entry)
XX DE Human hSI protein sequence.
XX KW Gastro-intestinal transport receptor; binding protein; hSI; HPT1; D2H;
KW hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;
KW intestinal peptide-associated transporter; hypertension; diabetes;
KW osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
XX KW therapeutic agent delivery; therapy; probe.
OS Homo sapiens.

XX	W09851325-A2.	
PN		
XX	19-NOV-1998.	
PD		
XX	15-MAY-1998;	98NO-US010088.
XX	15-MAY-1997;	97US-0046595P.
PR		
XX	(CYTO-) CYTOGEN CORP.	
XX	(ELAN-) ELAN CORP PLC.	
PA		
PA		
XX	Alvarez VL, Omahony DJ, Lambkin IJ, Patterson CA, Singleton J;	
PI	Belinka BA, Carter JM, Cagney GM;	
PI	WPI; 1999-009568/01.	
DR		
XX	New proteins that bind specifically to receptors in the gastro-intestinal	
PT	tract and related nucleic acid - chimeraas and antibodies, used to	
PT	deliver therapeutic or diagnostic agents to, or through, the	
PT	gastrointestinal tract, e.g. insulin or leuprolide.	
XX		
PS	Disclosure; Fig 3; 294pp; English.	
XX		
CC	This sequence is the human hSI protein. The invention relates to purified	
CC	proteins (I) that bind specifically to at least one of the gastro-	
CC	intestinal (GI) tract receptors human intestinal peptide-associated	
CC	transporter (hPT1), hPEP1, D2H and human sucrose-isomaltase complex	
CC	(hSI). (I) provide active transport of therapeutic agents through human	
CC	and animal GI tissue (into the blood) for in vivo delivery, particularly	
CC	for treatment or prevention of hypertension, diabetes, osteoporosis,	
CC	haemophilia, anaemia, cancer, migraine, or angina pectoris. Specifically	
CC	they are used to deliver insulin or leuprolide, but many other suitable	
CC	therapeutic agents are disclosed, including genes or inhibitory nucleic	
CC	acid, imaging agents and antigens. (I) may also provide targeting to the	
CC	GI tract. Other uses of (I) are: (i) to determine the level of specified	
CC	receptors in a sample (in a binding assay); and (ii) to screen for	
CC	molecules that bind (I). Immunogenic analogues or derivatives of (I) are	
CC	used to raise antibodies and in immunoassays. The antibodies are used	
CC	to locate, detect and measure (I), e.g. for imaging, monitoring treatment,	
CC	tissue analysis etc., also for peptide purification and immobilisation.	
XX		
SQ	Sequence 1827 AA;	
	Query Match 28.8%; Score 1877.5; DB 2; Length 1827;	
	Best Local Similarity 41.7%; Pred. No. 2.4e-149;	
	Matches 382; Conservative 157; Mismatches 317; Indels 59; Gaps 19;	
QY	340 AEATGAHGRPRRAVPTQCDVPPNSRFDCAPKAITQEOCEARGCCYIPAKOGLQAQMG 399	
DB	49 ATTRVTNPSDSGKCPNVLNDPVNVRINCIPEQPTTEGICAQRCGCCWRPNDSL----- 102	
QY	400 QPKCFPPSPSYKLENLSSESGYATLTR-TTPTFFPKDILTLRLDVMVMTENRLHFT 458	
DB	103 IPWCFFVDNH-GYNVODMTTISGVEAKLNRPSPPLTFGNDINSVLFTQNTQPNRFRPK 161	
QY	459 IKDPANRRYEVPLETERVHS--RAPSPLYSVFSEPEFPGVIVHRQLDGRVLNNTTVAPLF 516	
DB	162 ITDPNRRYEVPHQYKEFTGPTVSTLDYDKVAQNPFQIVIRKSNKTLFDTSGPLV 221	
QY	517 FADQFLQLSTSLPSQYITGLAEHL-SPLMLSTSWTRITLWNRDLAP-TFCANLYGSHPHY 574	
DB	222 YSDQYLQISARLPSDIYIGEGQVHKRFRHDLRSWKTWPIFTRDQLPGDNNNNLYGHQTF 281	
QY	575 LALED-GGSNAGVFLANSMDVVLOSPALSRSTGGILDVYIFGPPEKSVVQOYLDV 633	
DB	282 MCIEDTSGKSGFVFLMNSNAMEIFIQTPITVTRVGTGGILDFYILLGDTPEQVVQOYQQL 341	
QY	634 VGYFPMPYWGFLGFLHLCRCWYSSTALTROVVENNTRAHPFLDVQVMDLDYMSRRDFTFN 693	
DB	342 VGLPAMPAYNGLGFQSRWNYSKLDVVKEVVRNREAGIPFTQVTDIDYMEDKQDFTYD 401	
QY	694 KQGRDFPAMVQELHOGGRYMMIVDPAISSSGPAG--SYRPYDEGLRRGVFITNETGQ- 750	

DB	402 QVAFNGLPQFQVODLHDHGGKYVILDPALISIGRRANGTTYATYERGNTQHWVINSDSGDT 461	
QY	751 PLIGKVPFGSTAFDPDTNPTALAWMEDVAFPHDQVPFDGLMIDMNEPSNFIRSGEDGCP 810	
DB	462 PIIGEVWGLTYVPDFTNPNCIDWANECSIFHQEVQYDGLWIDWNEVSSFIQSGTKGCN 521	
QY	811 NNELENPPYVPGVGGTLOAAATICASSHQFTSTHYNLNLYGLTETAIASHRALVKA-RGT 869	
DB	522 VNKNYPPTDILDKLMYSKTCICMDAVQNMKGQVDVHSLYGYSMAIATEQAVQKVPFNK 581	
QY	870 RPFVTSRSTFAGHGRYAGHTWGDWSSWEOLASSVPEILQENLLGVPLVGVADVCPLGNT 929	
DB	582 RSFILTRSTFAGSGRHAHAWLGDNTASWEQMEWSITGMLEFSLFGIPLVGDICGFVAET 641	
QY	930 SEELCVRTQLCAFYPPFMRNHNLSLSPQEPYSFSEPA-----QOAMRKALTLRYAL 981	
DB	642 TEELCRRWQLGAFYPPFRNHN-----DGYEHQDPAFFGQNSLLVKSSRQYLIRYTL 695	
QY	982 LPHLYTLFQHAHVAGETVARPLFLFPKDSSTWTVHOLLMGEALLITPVLQAGKAEVTG 1041	
DB	696 LPFLYTLFVKAHVFGETVARPVLHEFYEDTNSWIEDTFLMGPALLITPVLKQGADTVSA 755	
QY	1042 YPPLGTWDLQTVPEALGSLUPPPAPAPREPAIHSEGOVTLPAPLDTINVHLRAGYIIP 1101	
DB	756 YIPDAIWDYES-----GAKRP-----WRKQKQVDMYLPADKIGLHLRGYIIP 798	
QY	1102 LQPGGLTTTREGROQPMALAVALTGKGEGARGLFWDGSELEVLERGAYTVQVIFLARNNTI 1161	
DB	799 IQEPDVTTASRKNPLGLIIVALGENTAKGDFWDDGETKOTIQNGVILYTFYSNNTL 858	
QY	1162 VNELVRVTS---EGAGLQKQVTVLGV--ATAPOQVLSNGVPV---SNFTYSPDTKVLDI 1213	
DB	859 --DIVCTHSSYQEGTTLAFQTVKILGLTSDVTEVRVAENQPMNAHSNFTYDASNQVLLI 916	
QY	1214 C-VSLLMGEQFLVSW 1227	
DB	917 ADLKNLGRNFSVQW 931	
	RESULT 13	
AAU09028	ID AAU09028 standard; protein; 1827 AA.	
XX	AC AAU09028;	
XX	DT 18-DEC-2001 (first entry)	
XX	DE Human sucrase isomaltase, SI.	
XX	KW Human; sucrase isomaltase; SI; stomach cancer; colorectal cancer;	
XX	KW oesophageal cancer; cytostatic; metastasis; CDX1; CDX2; tumour.	
OS	Homo sapiens.	
XX	FN WO200173133-A1.	
XX	PD 04-OCT-2001.	
XX	PF 27-MAR-2001; 2001WO-US009918.	
XX	PR 27-MAR-2000; 2000US-0192229P.	
XX	PA (UYJE-) UNIV JEFFERSON THOMAS.	
PI	Waldman SA, Park J, Schulz S;	
XX	WPI; 2001-611641/70.	
DR	N-PSDB; AAS14674.	
XX	PT In vitro screening for specific gastrointestinal cancer cells, useful for	
XX	diagnosis, by detecting expression of the markers SI, CDX1 or CDX2.	

PS Disclosure; Page 114-119; 119pp; English.

CC The invention relates to in vitro screening of metastatic colorectal cancer cells or primary and/or metastatic stomach or oesophageal cancer cells by testing cells in extra-intestinal tissues and/or body fluids for expression of SI (sucrase isomaltase), CDX1 or CDX2 (transcription factors). Expression of these markers indicates possible presence of the specified cancer cells. The method is used to diagnose (or monitor) metastatic colorectal cancer or primary and/or metastatic stomach or oesophageal cancer cells, also to confirm identification of such cells. These cancers can be treated by administration of an SI ligand and (optionally conjugated) cytostatic agent or radioimaged by administering a conjugate of the SI ligand and detectable agent. The present sequence represents human SI

XX

SQ Sequence 1827 AA;

Query Match 28.8%; Score 1877.5; DB 4; Length 1827;

Best Local Similarity 41.7%; Pred. No. 2.4e-149;

Matches 382; Conservative 157; Mismatches 317; Indels 59; Gaps 19;

QY 340 AEAEAGHGRPRAVPTQCDVPNPSRFDCAFDKAITQEQCEARGCCVPAKQGLQGAQM 399

DB 49 ATTRVTNPSDSKCPNVLNDPVNRINCIPQEPTEGICAQGCCWRPNWDSL----- 102

QY 400 QPWCFFPPSPYKLENLSSEMGYATLTR-TTPTFFPKDILTLRLDVMMETENLHFT 458

DB 103 IPWCFFVDNH-GYNVQDMTTTSGVEAKLRIPSTLFGNDINSVLTFTQNTQPNRFRK 161

QY 459 IKDPANRRVEVPLETRVHS--RAPPLYSVESEEPFVIVHRQLDGRVLLNTVAPLF 516

DB 162 ITDPNRRVEVPHQYKFGTPTVSTLYDVKVAQNPFSQIVIRKNGKTLFDTSGPLV 221

QY 517 FADQFLQLSTLPSQYITGLAEHL-SPLMLSTGWRITLWNRDLAP--TPCANLYGSHPFY 574

DB 222 YSDQYLIQISARLSDIYIGEQVHKRFRHDLSEKWTPIFTRQLPQDNNNLYGHQTFP 281

QY 575 LALED--GGSAGHGVFLNNSAMDVVLQSPALSWRSTGGILDVYIFLGPPEKSVVQYLDV 633

DB 282 MCIEDTSGKSGFVFLMNSNAMEIFIQTPITVYRVGTGGILDFYLLGDTPEQVVQYQQL 341

QY 634 VGYPPFPYKGLFHLRCWGYSSAITRQVVENWNTRAHPDLVDQNDLDMDSRDTFN 693

DB 342 VGLPAMPAYNNVGLFQSRMNYKSLDVVKEVVRNREAGIPFDQVTDIDMEDKDYTD 401

QY 694 KDGFRDPPAMVQSLHOGGRYMIIDVPAISSGPG--SYRPPYDEGLRGVFTTNETGQ- 750

DB 402 QVAFNGLPQVQDLHDHGQKXVILDPALISIGRANGTYATYERGNTQHVWLNESDGT 461

QY 751 PLIGKVPGSTAPDFTNPTALAWEDMVAEFDQVFPDGLWIDMNEPSNFIKGSBDGCP 810

DB 462 PLIGVMPGLTVYPDFTNPNCIDWANECSIFHQEVQYDGLWIDMNEVSFSIQSTKGCN 521

QY 811 NNELENPVYGVVGTQLQATTCASHQFSLTHYNLHLYGTEATASHRALVKA-RGT 869

DB 522 VNKXNLPPTPDTLKMLYKTKICMDPAQVQMGQYDVHSLYGSMAIATEQAVQKVPFNK 581

QY 870 RPPVSRSTFAGHGRVAGHTGVDWSSQELASSVPEILQNLGVLPLVCAQVCGFLGNT 929

DB 582 RSFILTRSTPAGSRHAHWLGONTASQEWESITGMLFSLFGLPLVGADICGVFAET 641

QY 930 SEELCVRWTLQGLAFYFMRNHNLSLIPQEPYFSFSPA-----QOAMKRLTLRYAL 981

DB 642 TEELCRWMLQGLAFYFMRNHNLS-----DGEHQDPAPFGQNSLLVKSRQVLTIRYL 695

QY 982 LPHLYTLFHOAHVAGETVARPLFEPKDSSTVVDHLLWGALLITPVLQKAEVTC 1041

DB 696 LPFLYTLFKARHVGETVARPLFHEFVEDTNSIEDTEFLWGALLITPVLKQADTVSA 755

QY 1042 YPLGTYDYLQTVPIEALGSLPAPPAAPREPAHSEGWTLTAPLDTTNVHLRAGYIP 1101

DB 756 YIPDAIWDYDES-----GAKRP-----WRKQRVDMYLPADKIGLHURGGYIP 798

QY 1102 LQPGGLTTTTSRQOPMALAVALTKGARGELFWDDGESLEVLERCAYTOVIFLARNTI 1161

DB 799 IOEPDVTTASRKNPLGLLVALGENTAKGDFWDDGETKDTIQNGYILYTFVSNNIL 858

QY 1162 VNELVRVTS---EGAGLQKQVTVLGV--ATAPOQVLSNGVPV---SNFTYSPDTKVLDI 1213

DB 859 --DIVCTHSSYQEGTTLAQTKVILGLTSDVTEVRVAENQPMNAHSNFTYDASQVLLI 916

QY 1214 C-VSLMLGEOFLVSW 1227

DB 917 ADLKLNLGRNFSVQW 931

RESULT 14

ABGI4994

ID ABGI4994 standard; protein; 1829 AA.

XX

AC ABGI4994;

XX

DT 18-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #14985.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US008631.

XX

PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YT;

XX

DR WPI; 2001-639362/73.

DR N-PSDB; AAS79181.

XX

PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity.

XX

PS Claim 20; SEQ ID NO 45353; 103pp; English.

XX

CC The invention relates to isolated polynucleotide (I) and polypeptide (II)

CC sequences. (I) is useful as hybridisation probes, polymerase chain

CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,

CC and in recombinant production of (II). The polynucleotides are also used

CC in diagnostics as expressed sequence tags for identifying expressed

CC genes. (I) is useful in gene therapy techniques to restore normal

CC activity of (II) or to treat disease states involving (II). (II) is

CC useful for generating antibodies against it, detecting or quantitating a

CC polypeptide in tissue, as molecular weight markers and as a food

CC supplement. (II) and its binding partners are useful in medical imaging

CC of sites expressing (II). (I) and (II) are useful for treating disorders

CC involving aberrant protein expression or biological activity. The

CC polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic

CC amino acid sequences of the invention. Note: The sequence data for this

CC patent did not appear in the printed specification, but was obtained in

CC electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences

XX

Query Match	28.8%	Score 1877.5;	DB 4;	Length 1829;
Best Local Similarity	41.7%;	Pred. No. 2.4e-149;		
Matches 382;	Conservative 157;	Mismatched 157;	Indels 59;	Gaps 19;

QY	340	A E A T G A H P R P R A V P P Q C D V P N S R F D C A P D K A I T Q E O C E A R G C C Y I P A K Q G L O G M Q G	399
DB	51	A T T R V T T N P S D S G K C P N V L A D P V N V R I N C I P E O P F T E G I C A Q R G C C W R P W N D S L - - - - -	104
QY	400	Q P W C F F P P S Y P S Y K L E N L S S E M G Y A T L A R - T P T F F P K D I L T L R L D V M M E T E N R L H F T	458
DB	105	I P W C F F V D N H - G Y N V Q D M T T S I G V E A K L N R I P S P T L F G N D I N S V L F T T Q N O P N R R F X	163
QY	459	I K D P A N R R Y E V P L E T P R V H S - R A P S P L Y S V F S E B P F G V I V H R Q L D G R V L L N T V T A P L F	516
DB	164	I T D P N N R R Y E V P H Q V Y K E F T G P T V S D T L Y D V K V A Q N P F S I Q V I R K S N G K T L F D T S I G P L V	223
QY	517	F A D Q F L O L S T S L S Q Y I T G L A E H L - S P L M L S T W T R I T L A N R L A P - T P G A N L Y G S H P F Y	574
DB	224	Y S Q Y I Q I S A R L P S D Y I G I G E Q V H K R F R U L S W K T W P I F T R Q L P G D N N N N I Y H G H T F F	283
QY	575	L A L E D - G G S A H G V F L L N S A M D V L Q P S P A L S W R S T G G I L D V I F L G P E K S V V Q Q L D V	633
DB	284	M C I E D T S G K S F G V F L M N S M A E I F Q P T I V T Y R V T G G I L D F Y I L L G D T P E Q V V Q Q Y Q L	343
QY	634	V G Y P F M P P Y K G L G F H L C R W G Y S T A I T R O V V E N M T R A H F P L D V Q W N D L D Y M D S R R D F T F N	693
DB	344	V G L P A M P A Y W N L G F O L S R W N Y K S L D V V K E V V R R N R E A G I F D T Q V T D I D Y M D K D F T Y D	403
QY	694	K D C F R D P P A M V Q E L H Q C R R Y M I V D P A I S S S G P A G - - S Y R P Y D E G L R R G V F T I N E T G Q -	750
DB	404	Q V A F N G L P Q F V Q D L H D G Q K V I L D P A I S G R A N G T T Y A T V E R G N T Q H V N E S D G S T	463
QY	751	P L I G K V W P G S T A P P D F T N P T A L A W E D M V A E F H D Q V P F D G L W I D M N S P N I F R G S E D G C P	810
DB	464	P I I G E V W P L T V Y P D F T P N C I D W A N E C S I F H Q E V Q Y D G L W I D M N S V S F I Q G S T K G C N	523
QY	811	N N E L E N P P Y P G V V G G T L Q A A T C A S H O F L S T H Y N L N L Y G L T E A T A S H R A L V K A - R G T	869
DB	524	V N K L N P P F P D I L D K L M Y S K T I C M D A V Q N W G K Q Y V H S L Y G Y S M A T A T S Q A Q K V P P N K	583
QY	870	R P F V I R S T P A G H R Y A G H W T G D V W S S W E O L A S S V P E I O F N L I G V P L V G A D V C G F L G N T	929
DB	584	R S F I L T R S T F A G S R H A A H L G N T A S W E Q M E N S I T G M L E F S L F G I P L V G A D I C G F V A E T	643
QY	930	S E E L C V R W T O L G A F Y P P M R N H N S L L S U P Q P Y S F S E P A - - - - - Q Q A W R K A L T L R Y A L	981
DB	644	T E E L C R R W M O L G A F Y P P S R N H S - - - - - D G Y E H Q D P A F F Q N S L L V K S R Q V L T I R Y T L	697
QY	982	L P H Y L T L F H O A H V A G E T V A R P L F L E P K D S S T W T V D H L W G E A L L I T P V L O A G K A B V T G	1041
DB	698	L P F Y L T L F Y K A H V F G E T V A R P V L H E F Y E D T N S W I E D T E F L W G P A L L I T P V L K Q G A D T V S A	757
QY	1042	Y F P L G T W Y D I L Q T V P I E A L G S L P P P P A P R E P A I H S E Q Q W T T L P A P L D T I N V H L R A G Y I I P	1101
DB	758	Y I P D A I W Y D E S - - - - - G A K R P - - - - - W R K Q R V D M Y L P A D K I G L H L R G G Y I I P	800
QY	1102	L Q S P G L T T T S R Q O P M A L A V A L T K G B A R G E L F W D D G E S L E V I E R G A Y T Q V I F L A R N T I	1161
DB	801	I Q E P D V T T A S R K N P L G L I V A L G E N N T A K G D F W M D G E T K D T I Q N G N Y I L Y T F S V S N N T L	860
QY	1162	V N E L V R V T S - - E G A G H Q L O K V T V L G V - - A T A P O O V L S N G V P V - - - S N F T Y S P D T K V L D I	1213
DB	861	--D I V C T H S S Y Q E G T T L A F Q T V K I L G I T D S V T E V R V A E N N Q P M A H S N F Y T D A S N Q V L L I	918
QY	1214	C - V S L L M G E O F L V S W	1227
DB	919	A D L K A L N L G R N F S V Q W	933

RESULT 15
AAB48844

XX	AA048844	standard; protein; 357 AA.
XX	AA048844;	
XX	AA048844;	
XX	13-MAR-2001	(first entry)
XX	Human RAP	(receptor associated protein).
XX	RAP; receptor	associated protein; human; factor VIII Clearance;
XX	LRP-mediated	plasma clearance; receptor-dependent Clearance;
XX	receptor-independent	clearance; ligand internalisation;
XX	low density	lipoprotein related protein; haemophilia; half-life.
XX	Homo sapiens.	
XX	WO200071714-A2.	
XX	30-NOV-2000.	
XX	24-MAY-2000;	2000WO-US014111.
XX	24-MAY-1999;	99US-0135847P.
XX	(AMNA-) AMERICAN	NAT RED CROSS.
XX	Saenko EL,	Strickland DK;
XX	WPI; 2001-025163/03.	
XX	N-PSDB; AAC48844.	
XX	Factor VIII	mutants having increased half-life useful for treating
XX	haemophilia,	comprise one or more amino acid substitutions in the A2
XX	and/or C2	domain of factor VIII.
XX	Claim 59;	Fig 4; 121pp; English.
XX	The invention	relates to human factor VIII mutants comprising an amino
XX	acid substitution	at one or more positions in the A2 domain and/or an
XX	amino acid substitution	at one or more positions in the C2 domain. The
XX	invention also	encompasses a factor VIII mutant which lacks a B domain
XX	(AA048842). The	factor VIII mutants have an increased half-life in the
XX	bloodstream. The	A2 domain mutants exhibit reduced LRP-dependent
XX	(receptor-dependent)	clearance of factor VIII, while C2 domain mutants
XX	have reduced	receptor-independent clearance. The invention also relates
XX	to a method of	using RAP (receptor associated protein), a protein which
XX	inhibits LRP	(low density lipoprotein related protein)-mediated ligand
XX	internalisation,	to increase the half-life of factor VIII. The mutant
XX	factor VIII	proteins, and nucleotides encoding them, are useful for
XX	treating	haemophilia. RAP, LRP-binding RAP mutants or fragments, and
XX	nucleic acids	encoding them may also be used in the treatment of
XX	haemophilia,	in combination with a mutant factor VIII protein or
XX	the invention.	The invention provides means of increasing the half-life
XX	of factor VIII	by reducing its clearance from plasma. The present
XX	sequence	represents human RAP
XX	Sequence 357	AA;

Query Match	25.5%	Score 1660.5;	DB 4;	Length 357;
Best Local Similarity	95.1%;	Pred. No. 4.4e-132;		
Matches 327; Conservative	2;	Mismatches 10;	Indels 5;	Gaps 1;

Qy	1	MRGPSALWILAL-----RTVLGSYSREKNPKPSKRESGEFFRWEKLINLWEKAQRL	55
	:	:	:
Db	10	LRLGLPALLLLLLPLGWPWAASHGGYKSREKNPKPSKRESGEFFRWEKLINLWEKAQRL	69
Qy	56	HLPVPVLAELHADLKIQERDELAWKCLKLDGLDGEDGEKEARLIIRNLNVILAKYGLDGKKD	115
Db	70	HLPVPVLAELHADLKIQERDELAWKCLKLDGLDGEDGEKEARLIIRNLNVILAKYGLDGKKD	129
Qy	116	ARQVTNSISGTQEDGLDDPRLEKLWHKAKTSKGFSGEELDCKLWRFLHHKEKVHEYNVL	175
Db	130	ARQVTNSISGTQEDGLDDPRLEKLWHKAKTSKGFSGEELDCKLWRFLHHKEKVHEYNVL	189

Qy	176	LETLSRTEETHENVISPSDLSDIKSGVLHSRHTELKEKLSINOGDLRLRRVSHQGYSTE	235
Db	190	LETLSRTEETHENVISPSDLSDIKSGVLHSRHTELKEKLSINOGDLRLRRVSHQGYSTE	249
Qy	236	AEFEPRVIDLWDLAQSANLTDKELEAFREELKHFEAKIEKHNHYQKQLEIAHEKLRHAE	295
Db	250	AEFEPRVIDLWDLAQSANLTDKELEAFREELKHFEAKIEKHNHYQKQLEIAHEKLRHAE	309
Qy	296	SVGDGERVSRREKHALLEGRTKELGYTVKXHLQDLSGRISRAR	339
Db	310	SVGDGERVSRREKHALLEGRTKELGYTVKXHLQDLSGRISRAR	353

Search completed: April 5, 2005, 01:02:24
Job time : 140 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2005, 22:59:17 ; Search time 55 Seconds
(without alignments)
1666.710 Million cell updates/sec

Title: US-10-600-862A-7
Perfect score: 6522
Sequence: 1 MRGPSALWLLALRTVLGS.....KVLDCVSLMGEQFLVSWC 1228

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/prodata1/iaa/5A COMB.pep.*
2: /cgn2_6/prodata1/iaa/5B COMB.pep.*
3: /cgn2_6/prodata1/iaa/6A COMB.pep.*
4: /cgn2_6/prodata1/iaa/6B COMB.pep.*
5: /cgn2_6/prodata1/iaa/PCTUS COMB.pep.*
6: /cgn2_6/prodata1/iaa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4744.5	72.7	952	4	US-09-949-016-5894
2	4744.5	72.7	960	4	US-09-949-016-9527
3	1877.5	28.8	1827	4	US-09-443-780C-14
4	1877.5	28.8	1827	4	US-09-079-723-179
5	1650	25.3	323	1	US-07-992-827D-1
6	1650	25.3	323	1	US-08-216-593-1
7	1650	25.3	323	5	PCT-US93-12380-1
8	1403.5	21.5	877	1	US-08-430-925A-4
9	1322	20.3	960	4	US-09-248-796A-16360
10	1314.5	20.2	682	4	US-09-744-926-2
11	1083.5	16.6	1003	4	US-09-949-016-10627
12	1061	16.3	209	1	US-07-992-827D-1
13	1061	16.3	209	5	PCT-US93-12380-2
14	1034	16.2	209	1	US-08-216-593-2
15	906	13.9	914	3	US-09-437-054A-8
16	891.5	13.7	919	3	US-09-437-054A-17
17	852.5	13.1	729	4	US-09-248-796A-17121
18	786.5	12.1	956	4	US-09-270-767-44032
19	751	11.5	205	1	US-07-992-827D-5
20	746	11.4	203	1	US-08-216-593-5
21	746	11.4	203	5	PCT-US93-12380-5
22	729.5	11.2	938	4	US-08-697-843A-1
23	664.5	10.2	693	4	US-09-376-343-2
24	515	7.9	275	4	US-09-248-796A-14129
25	509.5	7.8	788	4	US-09-489-039A-12831
26	485	7.4	449	4	US-09-248-796A-14130
27	422.5	6.5	1066	2	US-08-633-770A-1

28	422.5	6.5	1066	4	US-09-280-197-5	Sequence 5, Appli
29	422	6.5	1070	2	US-08-633-770A-2	Sequence 2, Appli
30	422	6.5	1070	4	US-09-280-197-6	Sequence 6, Appli
31	417.5	6.4	776	4	US-09-489-039A-10778	Sequence 10778, A
32	354	5.4	697	4	US-09-134-000C-5916	Sequence 5916, A
33	341	5.2	737	4	US-09-583-110-4038	Sequence 4038, Ap
34	341	5.2	742	4	US-09-107-433-3626	Sequence 3626, Ap
35	320	4.9	1091	3	US-08-633-768A-2	Sequence 2, Appli
36	318.5	4.9	1091	4	US-09-280-197-2	Sequence 2, Appli
37	311.5	4.8	1088	4	US-09-280-197-1	Sequence 1, Appli
38	311	4.8	1799	4	US-09-134-000C-5178	Sequence 5178, Ap
39	310.5	4.8	1088	3	US-08-633-768A-1	Sequence 1, Appli
40	307.5	4.7	1092	4	US-09-275-608-3	Sequence 3, Appli
41	257	3.9	390	4	US-09-634-238-331	Sequence 331, App
42	249	3.8	570	4	US-09-275-608-4	Sequence 4, Appli
43	173	2.7	1505	4	US-09-538-092-1102	Sequence 1102, Ap
44	170	2.6	150	3	US-09-437-054A-10	Sequence 10, Appl
45	165.5	2.5	1270	4	US-09-538-092-1321	Sequence 1321, Ap

ALIGNMENTS

RESULT 1

US-09-949-016-5894
; Sequence 5894, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5894
; LENGTH: 952
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-5894

Query Match	72.7%	Score	4744.5	DB	4	Length	952
Best Local Similarity	97.2%	Pred. No.	0				
Matches	890	Conservative	3	Mismatches	18	Indels	5
Gaps	1						
QY	313	LGRTKEIGYTVKKHLQDL	SGRISARAEATGAHPR	AVPTQCDVPPNSR	FDCAPDK	372	
DB	42	LGSSPVLTEHPHQCGASR	GPGR-----DAQAHPR	AVPTQCDVPPNSR	FDCAPDK	96	
QY	373	AITQCCARGCCYIPAKQ	LOGAQMGQPCFPFPPSY	SKLENSSSEMGTATL	TRTT	432	
DB	97	AITQCCARGCCYIPAKQ	LOGAQMGQPCFPFPPSY	SKLENSSSEMGTATL	TRTT	156	
QY	433	PTFFPKDILTLRLDVM	ETENLHFTIKDPANRR	VEVPLETPRVHSR	APSPLYVFSSE	492	
DB	157	PTFFPKDILTLRLDVM	ETENLHFTIKDPANRR	VEVPLETPRVHSR	APSPLYVFSSE	216	
QY	493	PRGVVHQLDGRVLNT	VAPLFFADQFLQIST	SPSQYITGLAEHL	SPLMLTSWTRI	552	
DB	217	PRGVVHQLDGRVLNT	VAPLFFADQFLQIST	SPSQYITGLAEHL	SPLMLTSWTRI	276	
QY	553	TLWNRLDAPTPCANLY	GSHPPFLALEDDGSG	ANGVFLNNSNMDV	LQSPALSWRSTG	612	
DB	277	TLWNRLDAPTPCANLY	GSHPPFLALEDDGSG	ANGVFLNNSNMDV	LQSPALSWRSTG	336	
QY	613	LDVYIFLGPBPKSVQ	QVLDVVGVFMPYPW	GFLGFLHLCRWG	YSSTAITRQVV	ENMTRAHF	672

Db 337 LDVYIFLGPEPKSVVQOYLDVVGVPMPYPWGLGFHLCRWGYSSTAITRQVVENMTRAHF 396
QY 673 PLDVQWNLDDYMSRRDFTFNKDGFRDPFAMVQELHOGGRRYMMIVDPAISSSGPAGSYR 732
Db 397 PLDVQWNLDDYMSRRDFTFNKDGFRDPFAMVQELHOGGRRYMMIVDPAISSSGPAGSYR 456
QY 733 PYDEGLRRGVFITNETGQPLIGKWPFGSTAFDPFTNPTALAWMEDVVAEFHDOVDPDGLW 792
Db 457 PYDEGLRRGVFITNETGQPLIGKWPFGSTAFDPFTNPTALAWMEDVVAEFHDOVDPDGLW 516
QY 793 IDMNPSNFIKRGSDGCPNNELENPYPVGVGGTLQAATICASSHQFLSTHYNLHNLXG 852
Db 517 IDMNPSNFIKRGSDGCPNNELENPYPVGVGGTLQAATICASSHQFLSTHYNLHNLXG 576
QY 853 LTAETASHRALVKARGTRPPVISRSTFAGHGRYAGHWTGDMVSSWELQASSVPEILQFNL 912
Db 577 LTAETASHRALVKARGTRPPVISRSTFAGHGRYAGHWTGDMVSSWELQASSVPEILQFNL 636
QY 913 LGVPLVGADVCGFLGNTSEELCVRTQLGAFYPMRNHNSLLSLPOEPYSFSEPAQAQMR 972
Db 637 LGVPLVGADVCGFLGNTSEELCVRTQLGAFYPMRNHNSLLSLPOEPYSFSEPAQAQMR 596
QY 973 KALTLYALLPHLYTLFHQAHVAGETVARPLFLEFPKDSSTWTDHQLLWGEALLITPV 1032
Db 697 KALTLYALLPHLYTLFHQAHVAGETVARPLFLEFPKDSSTWTDHQLLWGEALLITPV 756
QY 1033 QAGKAEVTGYPLGTWYDLQTVPIEALGSLPPPPAAPREPAIHSEGOQWTLPAPLDTINV 1092
Db 757 QAGKAEVTGYPLGTWYDLQTVPIEALGSLPPPPAAPREPAIHSEGOQWTLPAPLDTINV 816
QY 1093 HLRAGYIIPLOGPGLTTTTSRQOPMALAVATKGGEARGELFWDGSELEVLERGAYTOV 1152
Db 817 HLRAGYIIPLOGPGLTTTTSRQOPMALAVATKGGEARGELFWDGSELEVLERGAYTOV 876
QY 1153 IFLARNNTIVNELVRVTSEGAGLQKQVTLGVATAPQOVLNSGVPSNFTYSPDTKVLD 1212
Db 877 IFLARNNTIVNELVRVTSEGAGLQKQVTLGVATAPQOVLNSGVPSNFTYSPDTKVLD 936
QY 1213 ICVSLLMGEQFLVSWC 1228
Db 937 ICVSLLMGEQFLVSWC 952

RESULT 2

US-09-949-016-9527
; Sequence 9527, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9527
; LENGTH: 960
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9527

Query Match 72.7%; Score 4744.5; DB 4; Length 960;
Best Local Similarity 97.2%; Pred. NO. 0;
Matches 890; Conservative 3; Mismatches 18; Indels 5; Gaps 1;

RESULT 3

US-09-443-780C-14
; Sequence 14, Application US/09443780C
; Patent No. 6699973
; GENERAL INFORMATION:
; APPLICANT: O'Mahony, Daniel J
; APPLICANT: Seveso, Michela
; TITLE OF INVENTION: Antibodies To Peptides That Target G1T Receptors And Related Mathc

QY 313 LEGRTEKELGYTKKHLQDLISGRISRARAEAEATCAHPRGPRAPVPTQCDVPPNSRFDCAPOK 372
Db 50 LSGSSPVLEETHPAHQOGASRPGPR-----DAQAHPRGPRAPVPTQCDVPPNSRFDCAPOK 104
QY 373 AITQOCEARGCCYIPAKOGLQGAQMGOPWCFFPPSPYSYKLENISSSMGYSATLITRTT 432
Db 105 AITQOCEARGCCYIPAKOGLQGAQMGOPWCFFPPSPYSYKLENISSSMGYSATLITRTT 164
QY 433 PTFPPKDIILTLRLDVMETENELHFTIKDPANRRYEVEPLETPRVHSRAPSPLYSVEFSSE 492
Db 165 PTFPPKDIILTLRLDVMETENELHFTIKDPANRRYEVEPLETPRVHSRAPSPLYSVEFSSE 224
QY 493 PFGVTVHRQLDGRVLLNTTVAFLFFADQFLQSLSTSLPSQYITGLAEHLSPLMLSTSWTRI 552
Db 225 PFGVTVHRQLDGRVLLNTTVAFLFFADQFLQSLSTSLPSQYITGLAEHLSPLMLSTSWTRI 284
QY 553 TLWNRDLAPTGANLYGSHHPFYLALEDGGSAGHVFLNNSNMDVVLQPSPALSWRSTGGI 612
Db 285 TLWNRDLAPTGANLYGSHHPFYLALEDGGSAGHVFLNNSNMDVVLQPSPALSWRSTGGI 344
QY 613 LDVYIFLGPEPKSVVQOYLDVVGVPMPYPWGLGFHLCRWGYSSTAITRQVVENMTRAHF 672
Db 345 LDVYIFLGPEPKSVVQOYLDVVGVPMPYPWGLGFHLCRWGYSSTAITRQVVENMTRAHF 404
QY 673 PLDVQWNLDDYMSRRDFTFNKDGFRDPFAMVQELHOGGRRYMMIVDPAISSSGPAGSYR 732
Db 405 PLDVQWNLDDYMSRRDFTFNKDGFRDPFAMVQELHOGGRRYMMIVDPAISSSGPAGSYR 464
QY 733 PYDEGLRRGVFITNETGQPLIGKWPFGSTAFDPFTNPTALAWMEDVVAEFHDOVDPDGLW 792
Db 465 PYDEGLRRGVFITNETGQPLIGKWPFGSTAFDPFTNPTALAWMEDVVAEFHDOVDPDGLW 524
QY 793 IDMNPSNFIKRGSDGCPNNELENPYPVGVGGTLQAATICASSHQFLSTHYNLHNLXG 852
Db 525 IDMNPSNFIKRGSDGCPNNELENPYPVGVGGTLQAATICASSHQFLSTHYNLHNLXG 584
QY 853 LTAETASHRALVKARGTRPPVISRSTFAGHGRYAGHWTGDMVSSWELQASSVPEILQFNL 912
Db 585 LTAETASHRALVKARGTRPPVISRSTFAGHGRYAGHWTGDMVSSWELQASSVPEILQFNL 644
QY 913 LGVPLVGADVCGFLGNTSEELCVRTQLGAFYPMRNHNSLLSLPOEPYSFSEPAQAQMR 972
Db 645 LGVPLVGADVCGFLGNTSEELCVRTQLGAFYPMRNHNSLLSLPOEPYSFSEPAQAQMR 704
QY 973 KALTLYALLPHLYTLFHQAHVAGETVARPLFLEFPKDSSTWTDHQLLWGEALLITPV 1032
Db 705 KALTLYALLPHLYTLFHQAHVAGETVARPLFLEFPKDSSTWTDHQLLWGEALLITPV 764
QY 1033 QAGKAEVTGYPLGTWYDLQTVPIEALGSLPPPPAAPREPAIHSEGOQWTLPAPLDTINV 1092
Db 765 QAGKAEVTGYPLGTWYDLQTVPIEALGSLPPPPAAPREPAIHSEGOQWTLPAPLDTINV 824
QY 1093 HLRAGYIIPLOGPGLTTTTSRQOPMALAVATKGGEARGELFWDGSELEVLERGAYTOV 1152
Db 825 HLRAGYIIPLOGPGLTTTTSRQOPMALAVATKGGEARGELFWDGSELEVLERGAYTOV 884
QY 1153 IFLARNNTIVNELVRVTSEGAGLQKQVTLGVATAPQOVLNSGVPSNFTYSPDTKVLD 1212
Db 885 IFLARNNTIVNELVRVTSEGAGLQKQVTLGVATAPQOVLNSGVPSNFTYSPDTKVLD 944
QY 1213 ICVSLLMGEQFLVSWC 1228
Db 945 ICVSLLMGEQFLVSWC 960


```

Db      799 IQBPDVTTTASRKNPGLI VALGENTAKGDFPDDGGETKDTTQNGNYILYTSVSNNTL 858
Qy      1162 VNELVRVTS---EGAGLQLOKQVTLGV--ATAFQQVLSNGVPV---SNFTYSPTDKVLDI 1213
Db      859 -DIVCTHSSYQSGTTLATQTVKILGLTDSVTEVRVAENNPQPMNAHSNFTYDASNQVLLI 916
Qy      1214 C-VSLLMGEOFLVSW 1227
Db      917 ADLKLNLGRNFSVQW 931

RESULT 4
US-09-079-723-179
; Sequence 179, Application US/09079723
; Patent No. 6703362
; GENERAL INFORMATION:
; APPLICANT: Alvarez, Vernon L.
; APPLICANT: O'Mahony, Daniel J.
; APPLICANT: Lambkin, Imelda J.
; APPLICANT: Singleton, Judith
; APPLICANT: Patterson, Catherine A.
; APPLICANT: Cagney, Gerard M.
; APPLICANT: Belinka, Benjamin A.
; APPLICANT: Carter, John M.
; TITLE OF INVENTION: RANDOM PEPTIDES THAT BIND TO GASTRO-
; TITLE OF INVENTION: INTESTINAL TRACT (GIT) TRANSPORT RECEPTORS AND RELATE
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/079,723
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mierock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-219
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 179:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1827 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-09-079-723-179

Query Match      28.8%; Score 1877.5; DB 4; Length 1827;
Best Local Similarity 41.7%; Pred. No. 3.1e-153;
Matches 382; Conservative 157; Mismatches 317; Indels 59; Gaps 19;

Qy      340 AEAETGAHGRPRAPVPTQCDVPNRFDFCAPKAITOECEARGCCYIPAKQLOGAQM 399
Db      49 ATTRVTNPSDSKCPNVLNPNVNRINCPISQFPTGICAQGCCRPWNDSL----- 102
Qy      400 QPWCFFPPSPYSKLENLSSEMGVTATLTR--TTTFFPKDILTLLDVMMETENRLHFT 458
Db      103 IWCFFVDNH-GYNQDMMTTISIGVEAKLNRIPSPTLFGNDINSVLFTTQNTQPNRFRFK 161
Qy      459 IKDPANRRYVEPLETPRVHS--RAPSPLYSVBFSEPFQVIVHRRQDGRVLNLTTPVAPLF 516

```

Db 162 ITDENRRVEPHQYKEFTGPTVSTLYDVKAQNPFSIQVIRKSNKTLFDTSGPLV 221
QY FADQFLQLSTLPSQVITGLAHL-SPLMLSTSWTRITLWNRDLAP-TPCANLYGSHPPY 574
Db 222 YSDQYLQISARLPSDIYIGIEQVHRFRHDLKSWKTPFTDRDLQPDNNNLYGHQTF 281
QY LALED-GGSAGHVFLLNSNAMVVLQSPALSWRSTGGILDVYIFLGPEPKSVVQQYLDV 633
Db 282 MCIEDTSGKSGFVFLMNSAMEIFIQTPITVTVRVYTGIGILDFYILLGDTPEQVQYQQL 341
QY 634 VGYPPMPYGLGFHLCRWCYSTALTQVVENWHTAHPPLDVQWNLDMYMSRRDFTFN 693
Db 342 VGLPAMPAYWNLGFQLSRMNYKSLDVVKEVRRNREAGIPFDTQVTDIDYMEDKQFTYD 401
QY 694 KCGFRDPFAMVQLHCGRRYMMVDPALSSSPAG--SVRPYDEGLRGVFTTNETGQ- 750
Db 402 QVAFNGLPQVQLDHHGKQYVILDPALISGRANGTYAYVYRGNTQHWINESDGT 461
QY 751 PLGKVPGSTAPDFTNPTALAWMDVAEPHDQVFPDGLMIDMNEPSNFIRGSDGCP 810
Db 462 PLIGVMPGLTVYDFTNPNPCIDWANECSIPHOEVQYDGLMIDMNEVSSFIQGSTKGCN 521
QY 811 NNELENPPYVGVVGGTQQAATCASHQFLSTHYLNLNLYGLTEAIAHRAHLVKA-RGT 869
Db 522 VNKLANYPFPTDILDKLMYSKTCMDAVQWQKQYDVHSLYGYSMIAATEQAQVQVFPNK 581
QY 870 RPFVIRSTFAGHGRVAGHTGVDWWSBOLASSVPEILQFNLLGVPLVGDVCGPLGNT 929
Db 582 RSPILTRSTFAGSRAAHLWGDNTASWQMEWSITGMEFLSFGIPLVGADICGFVAET 641
QY 930 SEELCVRTQLGAFYFPMRNHNSLLSLPOEPYSFSEPA-----QOAMRKALTLRYAL 981
Db 642 TEELCRNMQLGAFYFPMRNHNS-----DGEHQDPAFFGQNSLLVKSSQVLTIRYTL 695
QY 982 LPHLYTLFQAHVAGETVARPLFEPKDSSTWVDHQLLWGEALLITVILQAGKAEVGT 1041
Db 696 LPFLYTLFKAHVFGETVARPLVHEFYEDTNSMIEDTEFLMGPAALLITVILKQGDVSA 755
QY 1042 YFPLGTYDLOTPIEALGSLPPPPAPPAEPATHSEGQWTLPAPLDTTNVHLRAGYIIP 1101
Db 756 YIPDAIWDYES-----GAKRP-----WRKQVDMVLPADKIGLHLRGYIIP 798
QY 1102 LQSPGTLTTESROOPMALAVATKGEARGELFWDGDESLEVLERGAYTQVIFLARNTI 1161
Db 799 IQSPDVTITASRKNPLGLIVALCENNTAKGDFWDDGETKDTIQNGNYILYTFVSNNTL 858
QY 1162 VNELRVTS---EGAGLQLOKVTVLGV--ATAQOVLNPGVPV---SNFTYSPDTKVLDI 1213
Db 859 --DIVCTHSSYQEGTTLAFQTKILGTDSTVTEVRVAENNPQMNHNSFTYDASNOVLLI 916
QY 1214 C-VSLLMGEOFLVSW 1227
Db 917 ADLKNLGRNFSVQW 931

RESULT 5
US-07-992-827D-1
; Sequence 1, Application US/07992827D
; Patent No. 5474766
; GENERAL INFORMATION:
; APPLICANT: Schwartz, Alan L., Bu, Guojun
; TITLE OF INVENTION: Methods And Compositions For
; TITLE OF INVENTION: Inhibition Of Hepatic Clearance Of Tissue-Type
; TITLE OF INVENTION: Plasmalogen Activator
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bryan Cave
; STREET: 245 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10167-0034

Query Match 25.3%; Score 1650; DB 1; Length 323;
Best Local Similarity 100.0%; Pred. No. 8.7e-135;
Matches 319; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 YSREKQKPKPKRSSEGEFRMEKLNQWKAQRLHLPVRLAELHADLKIQERDELAWK 80
Db 1 YSREKQKPKPKRSSEGEFRMEKLNQWKAQRLHLPVRLAELHADLKIQERDELAWK 60
QY 81 KLKLDGLDDEKEARLIRNLNVLAKYGLDGKDKARQVTSNLSGTQEDGDDPRLEKL 140
Db 61 KLKLDGLDDEKEARLIRNLNVLAKYGLDGKDKARQVTSNLSGTQEDGDDPRLEKL 120
QY 141 WHKAKTSKFGSEELDKLWREFLHHKEKVEYNNVLETLSTRTEEHENVISPSDLSIKG 200
Db 121 WHKAKTSKFGSEELDKLWREFLHHKEKVEYNNVLETLSTRTEEHENVISPSDLSIKG 180
QY 201 SVLHSRHTLEKELASINQGLDRLRLRVSHQGYSTAEFEPRVIDLWLAQSANTDKEL 260
Db 181 SVLHSRHTLEKELASINQGLDRLRLRVSHQGYSTAEFEPRVIDLWLAQSANTDKEL 240
QY 261 EAFREELKHFRAKIEKHNHYQKLEIAHEKLRHAEVSGDGRVSRREKHALLEGRTKEL 320
Db 241 EAFREELKHFRAKIEKHNHYQKLEIAHEKLRHAEVSGDGRVSRREKHALLEGRTKEL 300
QY 321 GYTVKKHLQDLSGRISRAR 339
Db 301 GYTVKKHLQDLSGRISRAR 319

RESULT 6
US-08-216-593-1
; Sequence 1, Application US/08216593
; Patent No. 5650391
; GENERAL INFORMATION:
; APPLICANT: Broze, George J., Jr.
; APPLICANT: Schwartz, Alan L.
; APPLICANT: Warshawsky, Ilka
; TITLE OF INVENTION: Methods and Compositions for Inhibition
; TITLE OF INVENTION: of Hepatic Clearance of Tissue Factor Pathway Inhibitor
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dennis A. Bennett, G.D. Searle & Co.,
; ADDRESSEE: Corporate Patent Dept.
; STREET: P. O. Box 5110
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA

QY 826 -----GTLQAATCASS----- 837
DB 509 SDYPLGDFKSNASEWKSISEAAATKTITTTSSASTSIDGKNTLAPKGNINYPYAIN 568
QY 838 -----HOFILSTH----- 877
DB 569 NDOGDHD-LATHAISPNATHADGTVEYNIHYIIGLQERAIYEAALLEIHPNKPEFIIIGRS 627
QY 878 TFAGHRYAGHWTDGWNSSWEQIASSVEILOPNLLGVPLVADGCGFNGTSEELCVRW 937
DB 628 SFAGSGKYGHWGSDNYADYIMYFSTPOALSMLSGIPFPFVDACGNGTDMELCSRW 687
QY 938 TOLGAYFPWRNHNLSLLSQEYSSEPAQAQMRKALTRYALLPHLYTLFHOAHVAGE 997
DB 688 MOLASFPFYNHNLVAGIPOEYVW-EAMKATKTSINRYSLPYTYTLHESHVTGI 746
QY 998 TVARPLFLFPKDSSTWTVDHQLWGEALLITPVLOAGKAEVGYFP--LGTWYDLQTVR 1055
DB 747 PINRAFNWQFPYSKELAGVDVTFQFVGDAALLVTVLEPGVNHVTKGIFPGENAVYIDFT-- 804
QY 1056 IEALGSLPPPPAAPREPAIHSEGOVWTLPAPLDTINVHLRAGYIIPLOQPGLTSTESROQ 1115
DB 805 -----HKKQKFTAGKNETLAAPLGHIPLHIKGGNIIPQEPGYTTTESRKN 850
QY 1116 PMALVALTKGEGARGLFWDDGESLEVLERGAYTOVIFLARNTIWNELVRYTSEGAGL 1175
DB 851 PFGLLVALDABGTASGLYLDGSDV--EALY--VDFVASKNKL--ASVFGSYEVR 904
QY 1176 Q-LOKVTVLGVATAPQOVLNSGVPSNFTYSPDTKVLDDICVSLMG---EOPLVSW 1227
DB 905 OPLANVTILOVDSEPKKVLFNNTVSHVYENGAVYITDLEKFTKEGAFAEFSIQW 960

RESULT 10
US-09-744-926-2
; Sequence 2, Application US/09744926
; Patent No. 679458
; GENERAL INFORMATION:
; APPLICANT: FROBERG, Claus
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING AN ALPHA-GLUCOSIDASE, PLANTS WHICH
; TITLE OF INVENTION: SYNTHESIZE A MODIFIED STARCH, THE GENERATION OF THE PLANTS, THEIR
; TITLE OF INVENTION: MODIFIED STARCH
; FILE REFERENCE: 514413-3865
; CURRENT APPLICATION NUMBER: US/09/744,926
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/EP99/05536
; PRIOR FILING DATE: 1999-07-31
; PRIOR APPLICATION NUMBER: DE 198 36 097.5
; PRIOR FILING DATE: 1998-07-31
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 682
; TYPE: PRT
; ORGANISM: Solanum tuberosum
US-09-744-926-2

Query Match 20.2%; Score 1314.5; DB 4; Length 682;
Best Local Similarity 40.0%; Pred. No. 5.3e-105;
Matches 279; Conservative 118; Mismatches 221; Indels 79; Gaps 19;

QY 571 HPFYL-----ALEDG-----GSAHGVFLNSNMDVLOPSPALSWRSTGGILDVYIFL 619
DB 13 HPIQLHRPALHRGYSFRYPAGVSHGVLLSSNGMDIV-YTGDRISKYKIVGILIDLYFFA 71
QY 620 GREPKSVQOYLDVGVPPMPYWGILGFLHCRWYSTAITROVNMTRAFPLDVQWN 679
DB 72 GPSPEMVVDQYTLQIRPAAMPYWSFGHQRWGYKNIDVLEWVDSYAKSRIPLEVMWT 131
QY 680 DLDYMSDRDFTPNKOGFDFPA-----MVQELHQGRYMMIVDPAISSGPAGSYRYV 734
DB 132 DIDYMGFKDFTLDP---VNFPLERVIFFLKRLHQNDQKTVLIDVPGISIN---NTDYT 185

QY 735 DSEGLRGVITNETGOLPIKGVKWPSTAPPDFTNPTALAWEDMVAEFHDOVPPDGLWID 794
DB 186 RRGMEADVEIKRD-NMPYQGVVMPGNVYYPDPATEVFEWRNEIEKFDQLVFPDGLWLD 244
QY 795 MNEPSNFIKRGSDGCGNNLENPYPVGVVGGTL--QAATICASSHQFLST-HYNLHNL 851
DB 245 MNELSNFI--TGPPYPSSTFPDPPYKINNKGDLPLNYRTVTPATSTHFGDTMEYNVHNL 302
QY 852 GITEAIAASHRALVKARGTRPFVISRSTPAGHRYAGHWTDGWNSSWEQIASSVPEILOPN 911
DB 303 GLEGRATYSALVNVTKRPFILVRSTPLGSKRYTSHMTGDNAAATWDLAYSIPITLISFG 362
QY 912 ILGVLVADGVCGLGNTSEELCVRWTLQAGVPPWRNHNLSLLSQEYSSEPAQAQMR 971
DB 363 LFGIPVMVGADICGFSNTEELCRRMIQLGAYFPFARDHSAKDTTPOEILYSWDSVA-AA 421
QY 972 RKALTRYALLPHLYTLFHOAHVAGETVARPLFLFPKDSSTWTVDHQLWGEALLITPV 1031
DB 422 KKVGLRLYQLLPYFYLWYEAHIGTPIARPLFPFSPQDAKTFDISTQFLGKGVWISPI 481
QY 1032 LQAGKAEVGYPLGTWYDL---QTVPIEALGSLPPPPAAPREPAIHSEGOVWTLPAPL 1087
DB 482 LKQGATSVDAVFPAGNWFDFNYSRSVL-----NOGTYMTLDAPP 522
QY 1088 DFINVHLRAGYIIPLOQPGLTSTESROQPMALVALTKGEGARGLFWDDGESLEV-LER 1146
DB 523 DHINVHVRGNLVMQGEAMTTQAAQRTAFKLLVLLSSKNSTGELFVDDDDDEVQMGREG 582
QY 1147 GAYTOVIFLARNTIWNELVRYTSE-----GAGLOLQKVTVLG-----VAT 1187
DB 583 GRWTLVKF---NSNIIGNKIVKSEVNGRYALDQGLVLEKVTLLGFGENVRLGKSYELVG 639
QY 1188 APQOVLNSGVPSNFTYSPDTKVLDDIC-VSLLMGSEQF 1223
DB 640 SHQQ--GNTMKESLKQSCQFVTMEISGMSILIGKEF 674

RESULT 11
US-09-949-016-10627
; Sequence 10627, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10627
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10627

Query Match 16.6%; Score 1083.5; DB 4; Length 1003;
Best Local Similarity 44.7%; Pred. No. 1.4e-84;
Matches 219; Conservative 79; Mismatches 149; Indels 43; Gaps 10;

QY 756 VMGSGTAPDFTNPTALAWEDMVAEFHDOVPPDGLWIDMNEPSNFIKRGSDGCGNNLE 815
DB 1 VMPGLTVYPTFTNPNCIDWANECSIFHQEVQYDGLWIDMNEVSSFIQSGTGCNVNKL 60
QY 816 NPYPVGVVGGTLQAATICASSHQFLSTHYNLHNLGLTEATASHRALVKA-RGTRPFVI 874
DB 61 YPFPTDILDKLWYSKTIKMDAVQNGKQYDVHSLGYSMATATEQAVQKVPFNKRKSFIL 120

Db 61 LSDIKGSLVLSRHTLKEKLSRINQGLDRLRRVSHQGYSTEARFPRVLDLWDLAQSAN 120
QY 255 LTDKLEAFREELKHFEAKIEKNHYQKQLEIAHEKLRHAESVGDGERSRSREKHALLE 314
Db 121 LTDKLEAFREELKHFEAKIEKNHYQKQLEIAHEKLRHAESVGDGERSRSREKHALLE 180
QY 315 GRTELGYTVKXHLQDLSGRISAR 339
Db 181 GRTELGYTVKXHLQDLSGRISAR 205

RESULT 14
US-08-216-593-2
; Sequence 2, Application US/08216593
; Patent No. 5650391
; GENERAL INFORMATION:
; APPLICANT: Broze, George J., Jr.
; APPLICANT: Schwartz, Alan L.
; APPLICANT: Marshawsky, Ilka
; TITLE OF INVENTION: Methods and Compositions for Inhibition
; TITLE OF INVENTION: of Hepatic Clearance of Tissue Factor Pathway Inhibitor
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dennis A. Bennett, G.D. Searle & Co.,
; ADDRESSEE: Corporate Patent Dept.
; STREET: P. O. Box 5110
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60680
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/216,593
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Bennett, Dennis A.
; REGISTRATION NUMBER: 34,547
; REFERENCE/DOCKET NUMBER: C-2795
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (708)470-6501
; TELEFAX: (708)470-6881
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 209 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-216-593-2

Query Match 16.2%; Score 1054; DB 1; Length 209;
Best Local Similarity 99.5%; Pred. No. 3e-83;
Matches 204; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 135 PRLEKLWHKAKTSKSGBELDKLWREFLHHKKEVHYNNVLLTSLRTBEIHENVISPSD 194
Db 1 PRLEKLWHKAKTSKSGBELDKLWREFLHHKKEVHYNNVLLTSLRTBEIHENVISPSD 60

QY 195 LSDIKGSLVLSRHTLKEKLSRINQGLDRLRRVSHQGYSTEARFPRVLDLWDLAQSAN 254
Db 61 LSDIKGSLVLSRHTLKEKLSRINQGLDRLRRVSHQGYSTEARFPRVLDLWDLAQSAN 120

QY 255 LTDKLEAFREELKHFEAKIEKNHYQKQLEIAHEKLRHAESVGDGERSRSREKHALLE 314
Db 121 LTDKLEAFREELKHFEAKIEKNHYQKQLEIAHEKLRHAESVGDGERSRSREKHALLE 180

QY 315 GRTELGYTVKXHLQDLSGRISAR 339
Db 181 GRTELGYTVKXHLQDLSGRISAR 205

RESULT 15
US-09-437-054A-8
; Sequence 8, Application US/09437054A
; Patent No. 6316698
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Kinney, Anthony J.
; TITLE OF INVENTION: Plant Alpha-Glucosidase II Homologs
; FILE REFERENCE: BB1273 US NA
; CURRENT APPLICATION NUMBER: US/09/437,054A
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: 60/107,909
; PRIOR FILING DATE: 1998-No. 6316698ember-10
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 8
; LENGTH: 914
; TYPE: PRT
; ORGANISM: Glycine max
US-09-437-054A-8

Query Match 13.9%; Score 906; DB 3; Length 914;
Best Local Similarity 27.4%; Pred. No. 3.3e-69;
Matches 252; Conservative 131; Mismatches 340; Indels 196; Gaps 25;

QY 407 PSVPSYKLENLSSEMGGYTATLT-----RTTP-----TFPPKDLTLRLDVMETENR 454
Db 46 PGSSSLIATDVTISCHDLTAKLTTPKHSQSEYKPLLLTSLVSQGLRLKID----- 97

QY 455 LHFTIKDPA-----NRRYEV-----LETPRVHSRAPSLYSVFSE----- 491
Db 98 -----EDPSLSPKKRFEVDPVIVSEFPSTKLWLPKISSVVENGLSSVYLSDGHSVLRH 152

QY 492 EPGVIVRHQLDGRVLLNTTAPLF-----FADQFLQSLTSLP----- 529
Db 153 DPPELFIRDDSGDRVISLNSHDLDFEQLKHKSEDDNWEQPRSHTRRYPGPQISPD 212

QY 530 -----SOYITGLAEHLSPMLSTSMTRITLNRDLAPTGAN----- 566
Db 213 VSYFGADFYVGPERRAASLALK-----PTRGPNVDESEPYRLFNLDVFEYI 258

QY 567 -----LYGSHFPYLALEDGSAHGVFLNLSNAMDV-----VLQSPALS--W 606
Db 259 HDSPFGLYGIPEFVMSHGKARGSGFFMLNAAEQIDVLAPGWDAESGIALPSHRIDTFW 318

QY 607 RSTGGILDVYIFLGPEPKSVVOQLDVVGYPPWPPYMWGLGFHLCRWGYSTAITROOVEN 666
Db 319 MSEAGVVDFAFFIGNPDKVLRQYTAVTGTTPAMPQLFSIAYHQRWNYRDEEDVEHVDK 378

QY 667 MTRAHFPPLDVQWNDLDYMDSRDDFTFNKDGFRDPFAMVQELHGGRRYMMIIVDPAISSG 726
Db 379 FBELDIPYDVLMDIEHTDGKRYFTWDRALFPHPEMQRKLASKGRHMTIIVDPHIKRO- 437

QY 727 PAGESVRYDEGLRRGVFIETNETGQPLIGKVPGSTAFPDFTNPTALAWEDMVA---EFHD 784
Db 438 --ENPHLKEASQKGYVVKDASGNDPFGWCPCSSSYPTDLNPEIRSWADKFSYQSYEG 495

QY 785 QVPFDGLWIDMNEPSNFIRGSDGECENNELEPNPPYVGVVGTLOAATCASSHQSLSTH 844
Db 496 STPSLYINDMNEPSVF-NGPEVTMPROVTH-----YGGV-----EH 531

QY 845 YNLHNLGLTEAIAASHRALVKARG---TRPFVISTFAGHGRYAGHWGTGDVWSSWEQA 901
Db 532 RELHAYGYFFMANTANGLLK-RGENDRPFFVLRSALFAGSQRYGAVTGDNTADHDLR 590

QY 902 SSVPEILOQLNLGVPLVGADVCGFLGNTSEELCVRWTLQGFYFPFNRNNSLLSLPOEY 961
Db 591 VSIWVLTGLTGMFSFGADIGGFFCGNPEPELLVRWYQLGAYVYPPFRAHAHDTKREPW 650

QY 962 SFSEPAQAMRKALTLRYALLPHLYTLFHOAHVAGETVARPLFLEFPKSSSTWTVDHQLL 1021

```
Db      651  LFGERNTELKDAIHVRYALLPYFTLFRANTTGVVVRPLWMEFPFDEATFSNDETFM 710
Qy      1022 WGEALLITPVLOAGKAEVTGYFP-LGTWYDLOTVPIEALGSLPPPAPREPAPAHSEGQW 1080
Db      711  VGSILVQGIYTERAKHASVYLPKQSWYDLRT-----GAVYKGGVT 752
Qy      1081 VTLPAPLDTINVHLRAGYIIPLOQP-GLTTESRQOPMALAVALTKGGEARGELFWDDGE 1139
Db      753  HKLEVTESIPAFORAGTIIARKDRFRSSTQMANDPYTLVVALNSSQAAGELYIDGGS 812
Qy      1140 SLEVLERGAYTVIFLARNTIIVNELVRVTSEGAGLQ-----LQKTVVLGVATAPQOVL 1194
Db      813  SFNFLQ-GGYIHRRFIFSNGKLTSIDLAPASSKGRYPSPDAFIERIILGHAPSSKNAL- 870
Qy      1195 NGVPVSNFTYSPDTKVLDI 1213
Db      871  -----IEPSNQKVDI 880
```

Search completed: April 5, 2005, 01:08:46
Job time : 59 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2005, 01:05:12 ; Search time 1601 Seconds
(without alignments)
254.339 Million cell updates/sec

Title: US-10-600-862A-7
Perfect score: 6522
Sequence: 1 MRPGSGALWLLALRTVLGS.....KVLDCVSLMGGEQFLVSWC 1228

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1413372 seqs, 331592847 residues

Total number of hits satisfying chosen parameters: 1413372

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications AA:*
- 1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/prodata/1/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/prodata/1/pubpaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/prodata/1/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/prodata/1/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/prodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/prodata/1/pubpaa/US09A_PUBCOMB.pep.*
 - 10: /cgn2_6/prodata/1/pubpaa/US09B_PUBCOMB.pep.*
 - 11: /cgn2_6/prodata/1/pubpaa/US09C_PUBCOMB.pep.*
 - 12: /cgn2_6/prodata/1/pubpaa/US09_NEW_PUB.pep.*
 - 13: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pep.*
 - 14: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pep.*
 - 15: /cgn2_6/prodata/1/pubpaa/US10C_PUBCOMB.pep.*
 - 16: /cgn2_6/prodata/1/pubpaa/US10D_PUBCOMB.pep.*
 - 17: /cgn2_6/prodata/1/pubpaa/US10_NEW_PUB.pep.*
 - 18: /cgn2_6/prodata/1/pubpaa/US11_NEW_PUB.pep.*
 - 19: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pep.*
 - 20: /cgn2_6/prodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6522	100.0	1228	17 US-10-600-862A-7	Sequence 7, Appli
2	6522	100.0	1228	17 US-10-812-849-7	Sequence 7, Appli
3	4744.5	72.7	952	17 US-10-943-893-2	Sequence 2, Appli
4	2024	31.0	707	14 US-10-032-189-125	Sequence 125, App
5	1877.5	28.8	1827	9 US-09-819-247-2	Sequence 2, Appli
6	1877.5	28.8	1827	15 US-10-428-225-2	Sequence 2, Appli
7	1877.5	28.8	1827	16 US-10-751-699-14	Sequence 14, Appli
8	1783	27.3	972	17 US-10-600-862A-9	Sequence 9, Appli
9	1783	27.3	972	17 US-10-812-849-9	Sequence 9, Appli
10	1702	26.1	463	17 US-10-600-862A-11	Sequence 11, Appli
11	1702	26.1	463	17 US-10-812-849-11	Sequence 11, Appli
12	1660.5	25.5	357	17 US-10-600-862A-21	Sequence 21, Appli
13	1660.5	25.5	357	17 US-10-868-577A-45	Sequence 45, Appli

14	1660.5	25.5	357	17	US-10-868-577A-53	Sequence 53, Appli
15	1660.5	25.5	357	17	US-10-812-849-21	Sequence 21, Appli
16	1660.5	25.5	682	17	US-10-741-600-1267	Sequence 1267, Ap
17	1660.5	25.5	682	17	US-10-741-600-1268	Sequence 1268, Ap
18	1650	25.3	320	17	US-10-812-849-1	Sequence 1, Appli
19	1850	25.3	323	9	US-09-792-793A-31	Sequence 31, Appli
20	1850	25.3	323	15	US-10-375-209A-31	Sequence 31, Appli
21	1850	25.3	323	15	US-10-600-862A-1	Sequence 1, Appli
22	1522.5	23.3	902	13	US-10-043-418-4	Sequence 4, Appli
23	1518	23.3	886	16	US-10-437-963-191618	Sequence 191618,
24	1494.5	22.9	903	13	US-10-043-418-3	Sequence 3, Appli
25	1481	22.7	906	15	US-10-424-599-155772	Sequence 155772,
26	1475.5	22.6	932	16	US-10-437-963-155467	Sequence 155467,
27	1458.5	22.4	914	13	US-10-043-418-2	Sequence 2, Appli
28	1437.5	22.0	752	15	US-10-425-114-57817	Sequence 57817, A
29	1403.5	21.5	877	13	US-10-043-418-1	Sequence 1, Appli
30	1396	21.4	909	16	US-10-437-963-151932	Sequence 151932,
31	1353.5	20.8	1075	15	US-10-369-493-3792	Sequence 3792, Ap
32	1327.5	20.4	378	17	US-10-600-862A-22	Sequence 22, Appli
33	1327.5	20.4	378	17	US-10-812-849-22	Sequence 22, Appli
34	1325.5	20.3	357	17	US-10-600-862A-23	Sequence 23, Appli
35	1325.5	20.3	357	17	US-10-812-849-23	Sequence 23, Appli
36	1292.5	19.8	930	15	US-10-424-599-284885	Sequence 284885,
37	1263	19.4	919	15	US-10-369-493-5379	Sequence 5379, Ap
38	1226.5	18.8	304	16	US-10-408-765A-1913	Sequence 1913, Ap
39	1215	18.6	856	15	US-10-369-493-6567	Sequence 6567, Ap
40	1200	18.4	890	15	US-10-369-493-3115	Sequence 3115, Ap
41	1195.5	18.3	348	17	US-10-600-862A-24	Sequence 24, Appli
42	1195.5	18.3	348	17	US-10-812-849-24	Sequence 24, Appli
43	1166.5	17.9	762	15	US-10-369-493-4166	Sequence 4166, Ap
44	1146	17.6	520	15	US-10-425-114-42226	Sequence 42226, A
45	1061	16.3	209	17	US-10-600-862A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-10-600-862A-7
; Sequence 7, Application US/10600862A
; Publication No. US20050026823A1
; GENERAL INFORMATION:
; APPLICANT: Zankel et al.
; TITLE OF INVENTION: USE OF THE CHAPERONE RECEPTOR-ASSOCIATED PROTEIN (RAP) FOR THE
; TITLE OF INVENTION: DELIVERY OF THERAPEUTIC COMPOUNDS TO THE BRAIN AND OTHER TISSUES
; FILE REFERENCE: 30610/39383
; CURRENT APPLICATION NUMBER: US/10/600,862A
; CURRENT FILING DATE: 2003-06-20
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7
; LENGTH: 1228
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: RAP-GAA fusion sequence
US-10-600-862A-7

Query Match	100.0%	Score	6522	DB	17	Length	1228
Best Local Similarity	100.0%	Pred. No.	0	Mismatches	0	Indels	0
Matches	1228	Conservative	0	Matches	0	Gaps	0
QY	1	MRPGSGALWLLALRTVLGSYSREKNQPKPSKRESGEEFRMEKLNQNLWEKQRLHPPV	60				
Db	1	MRPGSGALWLLALRTVLGSYSREKNQPKPSKRESGEEFRMEKLNQNLWEKQRLHPPV	60				
QY	61	RLAELHADLKIQERDELAWKKLKDLGLDGDGSKEARLIRNLNVLAKYGLDGKKDARQVT	120				
Db	61	RLAELHADLKIQERDELAWKKLKDLGLDGDGSKEARLIRNLNVLAKYGLDGKKDARQVT	120				
QY	121	SNSLSGTQBDGLDPRLEKLMWKAKTSKFSGEELDKLWREFLHHKKEHYENVLLETLS	180				
Db	121	SNSLSGTQBDGLDPRLEKLMWKAKTSKFSGEELDKLWREFLHHKKEHYENVLLETLS	180				

```
181 RTEEIHENVISPSDLSIDKGSVLHSRHTLKEKLSINQGLDRLRRVSHQGYSTAEFEE 240
181 RTEEIHENVISPSDLSIDKGSVLHSRHTLKEKLSINQGLDRLRRVSHQGYSTAEFEE 240
241 PRVIDLWDLAQSANLTDKELEAFREELKHFEAKIEKHNHYQKLEIAHEKLRHAESVGDG 300
241 PRVIDLWDLAQSANLTDKELEAFREELKHFEAKIEKHNHYQKLEIAHEKLRHAESVGDG 300
301 ERVSRREKHALLEGRTKELGYTVKKHLQDLSGRISRARAETGAHPGRPRAPVPTQCDV 360
301 ERVSRREKHALLEGRTKELGYTVKKHLQDLSGRISRARAETGAHPGRPRAPVPTQCDV 360
361 PPNRSRDCAPDKAITQEOCEARGCCYIPAKQGLQGAQMGPWCFFPPSPSYKLENLSS 420
361 PPNRSRDCAPDKAITQEOCEARGCCYIPAKQGLQGAQMGPWCFFPPSPSYKLENLSS 420
421 EMGYTATLTRTPTFPFKDILTRLDMVMEENRHLFTIKDPANRRYEVPLETPRVHSRA 480
421 EMGYTATLTRTPTFPFKDILTRLDMVMEENRHLFTIKDPANRRYEVPLETPRVHSRA 480
481 PPSPLYSEFSEEPFGVIVHRQDGRVLLNTTVAFLFFADQFLQSLPSQYITGLAEHL 540
481 PPSPLYSEFSEEPFGVIVHRQDGRVLLNTTVAFLFFADQFLQSLPSQYITGLAEHL 540
541 SPLMLSTSWTRITLWNRDLAPTGANLYGSHHPFYLALEDGGAHGVFLNLSNAMDVLQP 600
541 SPLMLSTSWTRITLWNRDLAPTGANLYGSHHPFYLALEDGGAHGVFLNLSNAMDVLQP 600
601 SPALSWRSTGGILDVYIFLGPPEKSVVQYLDVVGYPFMPYVGLGFHLCRWGYSTAIT 660
601 SPALSWRSTGGILDVYIFLGPPEKSVVQYLDVVGYPFMPYVGLGFHLCRWGYSTAIT 660
661 ROVENNMTAHPFLDVQWMDLDYMDSRDFTFNKGFDRFPAMVQELHGGRRYMMIVDP 720
661 ROVENNMTAHPFLDVQWMDLDYMDSRDFTFNKGFDRFPAMVQELHGGRRYMMIVDP 720
721 AISSSGPAGSYRYPDEGLRGVPIINETGQPLIGKVPGSTAFPDPFTNPTALAWMDMVA 780
721 AISSSGPAGSYRYPDEGLRGVPIINETGQPLIGKVPGSTAFPDPFTNPTALAWMDMVA 780
781 EFHDQVFPDGLWIDMNEPNSFRGSDGCPNNELENPPYVGVGTTLQAATICASSHOF 840
781 EFHDQVFPDGLWIDMNEPNSFRGSDGCPNNELENPPYVGVGTTLQAATICASSHOF 840
841 LSTHYNLNLYGLTEAIAASHRALVKARGTRPPVISRSTFAGHRYAGHWTGDVWSSWEOL 900
841 LSTHYNLNLYGLTEAIAASHRALVKARGTRPPVISRSTFAGHRYAGHWTGDVWSSWEOL 900
901 ASSVPEILOFNLLGVLGVADVCGFLGNTSEELCVRWTLQGFAPYPMRNHNSLLSLPQEP 960
901 ASSVPEILOFNLLGVLGVADVCGFLGNTSEELCVRWTLQGFAPYPMRNHNSLLSLPQEP 960
961 YSFSEPAQAMRKALTLRYALLPHLYTLFQHAHVAGETVARPLFLFPKDSSTWVDHOL 1020
961 YSFSEPAQAMRKALTLRYALLPHLYTLFQHAHVAGETVARPLFLFPKDSSTWVDHOL 1020
1021 LMGEALLITPVLOAGKAEVTPPLGTWYDLOTVPIEALGSLPPPAAPREPAIHSEGOM 1080
1021 LMGEALLITPVLOAGKAEVTPPLGTWYDLOTVPIEALGSLPPPAAPREPAIHSEGOM 1080
1081 VTLPAPDLTINVHLRAGYIIPLOGPGLTITTESRQOPMALAVALTKEGARGELFWDGDS 1140
1081 VTLPAPDLTINVHLRAGYIIPLOGPGLTITTESRQOPMALAVALTKEGARGELFWDGDS 1140
1141 LEVLERGAYQVIFLARNNTINVELVRVTSBAGLQLOKVTVLGVATAPQQVLSNGVPVS 1200
1141 LEVLERGAYQVIFLARNNTINVELVRVTSBAGLQLOKVTVLGVATAPQQVLSNGVPVS 1200
1201 NFTYSPDTKVLDICVSLANGEOPLVSMC 1228
1201 NFTYSPDTKVLDICVSLANGEOPLVSMC 1228
```

```
RESULT 2
US-10-812-849-7
; Sequence 7, Application US/10812849
; Publication No. US20050042227A1
; GENERAL INFORMATION:
; APPLICANT: Zankel et al.
; TITLE OF INVENTION: MEGALIN-BASED DELIVERY OF THERAPEUTIC COMPOUNDS TO THE BRAIN AND
; OTHER TISSUES
; FILE REFERENCE: 30610/40037
; CURRENT APPLICATION NUMBER: US/10/812,849
; PRIOR FILING DATE: 2004-03-30
; PRIOR APPLICATION NUMBER: US 10/600,862
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7
; LENGTH: 1228
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: RAP-GAA fusion sequence
US-10-812-849-7
```

Query Match 100.0%; Score 6522; DB 17; Length 1228;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MRGPGALWLLIALRTVLGYSREKNQPKSPKRESGEFRMEKLNQLWEKAORLHLPV 60
DB 1 MRGPGALWLLIALRTVLGYSREKNQPKSPKRESGEFRMEKLNQLWEKAORLHLPV 60
QY 61 RLAEHLADLKIQRDELAWKKLKDGLDEDEGKEARLIRNLNVLAKYGLDGKKDARQVT 120
DB 61 RLAEHLADLKIQRDELAWKKLKDGLDEDEGKEARLIRNLNVLAKYGLDGKKDARQVT 120
QY 121 SNSLSGTQEDGLDDPRLEKLMHKAITSKFGSEELDKLWREFLHHKEKVHEYNVLETL 180
DB 121 SNSLSGTQEDGLDDPRLEKLMHKAITSKFGSEELDKLWREFLHHKEKVHEYNVLETL 180
QY 181 RTEETHENVISPSDLSIDKGSVLHSRHTLKEKLSINQGLDRLRRVSHQGYSTAEFEE 240
DB 181 RTEETHENVISPSDLSIDKGSVLHSRHTLKEKLSINQGLDRLRRVSHQGYSTAEFEE 240
QY 241 PRVIDLWDLAQSANLTDKELEAFREELKHFEAKIEKHNHYQKLEIAHEKLRHAESVGDG 300
DB 241 PRVIDLWDLAQSANLTDKELEAFREELKHFEAKIEKHNHYQKLEIAHEKLRHAESVGDG 300
QY 301 ERVSRREKHALLEGRTKELGYTVKKHLQDLSGRISRARAETGAHPGRPRAPVPTQCDV 360
DB 301 ERVSRREKHALLEGRTKELGYTVKKHLQDLSGRISRARAETGAHPGRPRAPVPTQCDV 360
QY 361 PPNRSRDCAPDKAITQEOCEARGCCYIPAKQGLQGAQMGPWCFFPPSPSYKLENLSS 420
DB 361 PPNRSRDCAPDKAITQEOCEARGCCYIPAKQGLQGAQMGPWCFFPPSPSYKLENLSS 420
QY 421 EMGYTATLTRTPTFPFKDILTRLDMVMEENRHLFTIKDPANRRYEVPLETPRVHSRA 480
DB 421 EMGYTATLTRTPTFPFKDILTRLDMVMEENRHLFTIKDPANRRYEVPLETPRVHSRA 480
QY 481 PPSPLYSEFSEEPFGVIVHRQDGRVLLNTTVAFLFFADQFLQSLPSQYITGLAEHL 540
DB 481 PPSPLYSEFSEEPFGVIVHRQDGRVLLNTTVAFLFFADQFLQSLPSQYITGLAEHL 540
QY 541 SPLMLSTSWTRITLWNRDLAPTGANLYGSHHPFYLALEDGGAHGVFLNLSNAMDVLQP 600
DB 541 SPLMLSTSWTRITLWNRDLAPTGANLYGSHHPFYLALEDGGAHGVFLNLSNAMDVLQP 600
QY 601 SPALSWRSTGGILDVYIFLGPPEKSVVQYLDVVGYPFMPYVGLGFHLCRWGYSTAIT 660
DB 601 SPALSWRSTGGILDVYIFLGPPEKSVVQYLDVVGYPFMPYVGLGFHLCRWGYSTAIT 660
QY 661 ROVENNMTAHPFLDVQWMDLDYMDSRDFTFNKGFDRFPAMVQELHGGRRYMMIVDP 720
```

Db 661 RQVENMTRAHFPLDVQMDLNDYMSRRDFTFNKDGFRDPPAMVQELHOGGRRYMMI VDP 720
Qy 721 A1SSSGPAGSYRYDEGLRGVITNETGOLPIKGVWPGSTAPDPDTNPTALAWEDMVA 780
Db 721 A1SSSGPAGSYRYDEGLRGVITNETGOLPIKGVWPGSTAPDPDTNPTALAWEDMVA 780
Qy 781 EFHQVPPFGLWIDMNEPSNFIKSGSDGCPNNELENPPYVGVGGTLOAATTCASSHQF 840
Db 781 EFHQVPPFGLWIDMNEPSNFIKSGSDGCPNNELENPPYVGVGGTLOAATTCASSHQF 840
Qy 841 LSTHYNLHNLGYTEALASHRALVKARGTRPFVSRSTFAGHGRYAGHWTGDVWSSWEQL 900
Db 841 LSTHYNLHNLGYTEALASHRALVKARGTRPFVSRSTFAGHGRYAGHWTGDVWSSWEQL 900
Qy 901 ASSVPBILQFNLLGVLGADVCGFLGNTSEELCVRWTLQGFYPPMRNHNLSLLSQEP 960
Db 901 ASSVPBILQFNLLGVLGADVCGFLGNTSEELCVRWTLQGFYPPMRNHNLSLLSQEP 960
Qy 961 YSPSEPAQAMRKALTRVALLPHLYTLFHOAHVAGETVARPLFLFPKDSSTWTVDHQL 1020
Db 961 YSPSEPAQAMRKALTRVALLPHLYTLFHOAHVAGETVARPLFLFPKDSSTWTVDHQL 1020
Qy 1021 LMGEALLITPVLOAGKAEVTGYPLGTWYDLOTVPIEALGSLPPPAPRPAIHSEGOV 1080
Db 1021 LMGEALLITPVLOAGKAEVTGYPLGTWYDLOTVPIEALGSLPPPAPRPAIHSEGOV 1080
Qy 1081 VTLPAPLDTINVLHLAGYIIPLOGPGLTTTESRQQPMALAVALTKEGARGELFWDGGS 1140
Db 1081 VTLPAPLDTINVLHLAGYIIPLOGPGLTTTESRQQPMALAVALTKEGARGELFWDGGS 1140
Qy 1141 LEVLERGAYTQVIFLARNNTIVNELVRVTSEGAGLQKQVTLGVATAPQOVLNSGVVPS 1200
Db 1141 LEVLERGAYTQVIFLARNNTIVNELVRVTSEGAGLQKQVTLGVATAPQOVLNSGVVPS 1200
Qy 1201 NFTYSPDTKVLDICVSLLMGEQFLVSWC 1228
Db 1201 NFTYSPDTKVLDICVSLLMGEQFLVSWC 1228

RESULT 3
US-10-943-893-2
; Sequence 2, Application US/10943893
; Publication No. US20050058634A1
; GENERAL INFORMATION:
; APPLICANT: Genzyme Corporation
; TITLE OF INVENTION: Methods for Introducing Mannose 6-Phosphate and Other Oligosaccha
; FILE REFERENCE: 07680.0031-01000
; CURRENT APPLICATION NUMBER: US/10/943,893
; CURRENT FILING DATE: 2004-09-20
; PRIOR APPLICATION NUMBER: US 10/051,711
; PRIOR FILING DATE: 2002-01-17
; PRIOR APPLICATION NUMBER: 60/263,078
; PRIOR FILING DATE: 2001-01-18
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 952
; TYPE: PRT
; ORGANISM: Artificial
; OTHER INFORMATION: recombinant human acid alpha glucosidase
; FEATURE:
; NAME/KEY: recombinant human acid alpha glucosidase
; LOCATION: (1)..(952)
; OTHER INFORMATION:
US-10-943-893-2

Query Match 72.7%; Score 4744.5; DB 17; Length 952;
Best Local Similarity 97.2%; Pred. No. 0;
Matches 890; Conservative 3; Mismatches 18; Indels 5; Gaps 1;

Qy 313 LBGRTKELGYTVKKHLQDLQSGRISRARAEATGAHGRPRAVPTQCDVPPNSRFDCAPOK 372
Db 42 LSGSSPVLBEETHPAHQOGASRPGPR-----DAQAHEGRPRAVPTQCDVPPNSRFDCAPOK 96
Qy 373 AITQEOCEARGCCYIPAKOGLQAGQWPCWPPPPSYPSYKLENLSSSBSMGYATATITRTT 432
Db 97 AITQEOCEARGCCYIPAKOGLQAGQWPCWPPPPSYPSYKLENLSSSBSMGYATATITRTT 156
Qy 433 PTFFPKDILTLRLDVMMETENLHFTIKDPANRRYEVPLETFRVHSRAPSPLYSVFSEE 492
Db 157 PTFFPKDILTLRLDVMMETENLHFTIKDPANRRYEVPLETFRVHSRAPSPLYSVFSEE 216
Qy 493 PFGVIVHROLDORVLLNTTVAFLPADQFLQSLPSQYITGLAEHLSPMLSTSWTRI 552
Db 217 PFGVIVHROLDORVLLNTTVAFLPADQFLQSLPSQYITGLAEHLSPMLSTSWTRI 276
Qy 553 TLWNDRLATPTGANLYGSHPPFYLLALEDDGSAUGVFLNNSNAMDVVLQPSPALSWRSTGGI 612
Db 277 TLWNDRLATPTGANLYGSHPPFYLLALEDDGSAUGVFLNNSNAMDVVLQPSPALSWRSTGGI 336
Qy 613 LDVYIFLGPPEKSVVQYLDVVGYPPMPYWGILGPHLCRWGYSSTAITRQVVENMTRAHF 672
Db 337 LDVYIFLGPPEKSVVQYLDVVGYPPMPYWGILGPHLCRWGYSSTAITRQVVENMTRAHF 396
Qy 673 PLDVQWNLQDYMDSRRDFTFNKDGFRDPPAMVQELHOGGRRYMMI VDPALSSSGPAGSYR 732
Db 397 PLDVQWNLQDYMDSRRDFTFNKDGFRDPPAMVQELHOGGRRYMMI VDPALSSSGPAGSYR 456
Qy 733 PYDEGLRRGVITNETGOLPIKGVWPGSTAPDPDTNPTALAWEDMVAHFDDVDPDGLW 792
Db 457 PYDEGLRRGVITNETGOLPIKGVWPGSTAPDPDTNPTALAWEDMVAHFDDVDPDGLW 516
Qy 793 IDWNEPSNFIKSGSDGCPNNELENPPYVGVGGTLOAATTCASSHQFLSTHYNLHNLGY 852
Db 517 IDWNEPSNFIKSGSDGCPNNELENPPYVGVGGTLOAATTCASSHQFLSTHYNLHNLGY 576
Qy 853 LFEATASHRALVKARGTRPFVSRSTFAGHGRYAGHWTGDVWSSWQQLASSVPEILQFNL 912
Db 577 LFEATASHRALVKARGTRPFVSRSTFAGHGRYAGHWTGDVWSSWQQLASSVPEILQFNL 636
Qy 913 LGVPLVGADVCGFLGNTSEELCVRWTLQGFYPPMRNHNLSLLSQEPYSFSEPAQAMR 972
Db 637 LGVPLVGADVCGFLGNTSEELCVRWTLQGFYPPMRNHNLSLLSQEPYSFSEPAQAMR 696
Qy 973 KALTLYALLPHLYTLFHOAHVAGETVARPLFLFPKDSSTWTVDHQLWGEALLITPVL 1032
Db 697 KALTLYALLPHLYTLFHOAHVAGETVARPLFLFPKDSSTWTVDHQLWGEALLITPVL 756
Qy 1033 QAGKAEVTGYPLGTWYDLOTVPIEALGSLPPPAPRPAIHSEGOVWTLPAPLDTINV 1092
Db 757 QAGKAEVTGYPLGTWYDLOTVPIEALGSLPPPAPRPAIHSEGOVWTLPAPLDTINV 816
Qy 1093 HLRAGYIIPLOGPGLTTTESRQQPMALAVALTKEGARGELFWDGGSLEVLERGAYTOV 1152
Db 817 HLRAGYIIPLOGPGLTTTESRQQPMALAVALTKEGARGELFWDGGSLEVLERGAYTOV 876
Qy 1153 IFLARNNTIVNELVRVTSEGAGLQKQVTLGVATAPQOVLNSGVVPSNFTYSPDTKVLD 1212
Db 877 IFLARNNTIVNELVRVTSEGAGLQKQVTLGVATAPQOVLNSGVVPSNFTYSPDTKVLD 936
Qy 1213 ICVSLLMGEQFLVSWC 1228
Db 937 ICVSLLMGEQFLVSWC 952

RESULT 4
US-10-032-189-125
; Sequence 125, Application US/10032189
; Publication No. US20030170630A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook II, John P
; APPLICANT: Tchernev, Velizar T

APPLICANT: Liu, Xiaohong
APPLICANT: Spytek, Kimberly A
APPLICANT: Zerhusen, Bryan D
APPLICANT: Patturajan, Meera
APPLICANT: Grosse, William M
APPLICANT: Lepley, Denise M
APPLICANT: Burgess, Catherine E
APPLICANT: Shimkets, Richard A
APPLICANT: Grosse, William M
APPLICANT: Szekeres, Edward S
APPLICANT: Vernet, Corine A.M.
APPLICANT: Li, Li
APPLICANT: Casman, Stacie J
APPLICANT: Boldog, Ferenc L
APPLICANT: Gorman, Linda
APPLICANT: Gangolli, Baha A
APPLICANT: Fernandes, Elma R
APPLICANT: Rieger, Daniel K
APPLICANT: Edinger, Shlomit R
APPLICANT: Gunther, Erik
APPLICANT: Millet, Isabelle
APPLICANT: Sciore, Paul
APPLICANT: Ellerman, Karen
APPLICANT: MacDougall, John R
APPLICANT: Smithson, Glennda
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-228
CURRENT APPLICATION NUMBER: US/10/032,189
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: 60/257,495
PRIOR FILING DATE: 2000-12-21
PRIOR APPLICATION NUMBER: 60/258,171
PRIOR FILING DATE: 2000-12-20
PRIOR APPLICATION NUMBER: 60/269,940
PRIOR FILING DATE: 2001-02-20
PRIOR APPLICATION NUMBER: 60/274,192
PRIOR FILING DATE: 2001-03-08
PRIOR APPLICATION NUMBER: 60/277,826
PRIOR FILING DATE: 2001-03-22
PRIOR APPLICATION NUMBER: 60/279,840
PRIOR FILING DATE: 2001-03-29
PRIOR APPLICATION NUMBER: 60/282,981
PRIOR FILING DATE: 2001-04-11
PRIOR APPLICATION NUMBER: 60/283,656
PRIOR FILING DATE: 2001-04-13
PRIOR APPLICATION NUMBER: 60/309,247
PRIOR FILING DATE: 2001-07-31
PRIOR APPLICATION NUMBER: 60/311,754
PRIOR FILING DATE: 2001-08-17
PRIOR APPLICATION NUMBER: 60/313,331
PRIOR FILING DATE: 2001-08-17
NUMBER OF SEQ ID NOS: 260
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 125
TYPE: PRT
LENGTH: 707
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Glycosyl
OTHER INFORMATION: hydrolases family 31

Query Match 31.0%; Score 2024; DB 14; Length 707;
Best Local Similarity 54.7%; Pred. No. 2.6e-149;
Matches 399; Conservative 87; Mismatches 213; Indels 30; Gaps 9;
QY 470 PIETPRVHGRAPSLYVSFSEPFVHVRQDGRVLLNTTVAPLFFADQFLQLSTSLP 529
Db 1 PLDFPASSPAESDLYDLVNSNGPFGFEVIRKSTGDLFDFTTFFGLVFSQDLQLSTSLP 60
QY 530 SQVITGLAHLSPMLSTWTRITLWNRDLATPPG-ANLYGSHPPFVLALEDGSGAHGVFL 588
Db 61 SEYIIGLGEHAHKLFRDRTNETVTLWNRDVGPSYGDNNLYGSHPPFYMSLEDSGNAHGVFL 120

QY 589 LNSNAMDVVLPSPALSWRSTGGILDVYIFLGPEKSVVQOYLDVVGVYFPMPPYMGSLGFH 648
Db 121 LNSNAMEVDIGPGPALTFRVIGGILDVFFFLGPTEDVLQOYTELIGRALPPYMSLGFH 180
QY 649 LCRWGYSTAITRQVVENWTRAHFPDLQVMDNDLWMSRRDFTFNKGRDPRDPAMVQOELH 708
Db 181 LCRWGYTNVSEVKTVVDGMRKANIPLDVQWLDIDYMDGYKDFTPVPRPFGPEDFVKLHL 240
QY 709 QGGRYMMIVDPAISSGPGAGSYRYPVDEGLRGVITNETGQPLICKVWPGSGTAPDFTN 768
Db 241 AKGQKIVVILDPAISVD--SASYPYERKKGKGVFKNPNGSDYIGEVWPGYTAFFDFTN 298
QY 769 PTALAWEDMVAEFHDQVPFDGLWIDMNEPSNFIKRGSDGCPNNELENPPYVGVVGGTL 828
Db 299 PEARKWADEIKDFHDSLPPFDGIWIDMNEPSSF---SEKGPNDNLNYPYAPNDGDGFL 355
QY 829 QAAITCASS-HQFLSTHYNLHLYGLTEAIAASHRALVKAR-CTRPFVISRSTFAGHGRYA 886
Db 356 SSKTMCMDAVHYGGVEHYDVHNLGLSEAKATYEALKKVTGGRPFVLSRSTFAGSGRYA 415
QY 887 GHWTGDMVMSWELASSVPEILQPNLGVPLVADVCGPLGNTSEELCVRWTLQGAFFP 946
Db 416 GHWTGENTASDLDLKYISIPGVLSFNLFGPIPFVGAIDICGPNGTTEELCVRMQLGAFFP 475
QY 947 MNHNSLLSLPOEYPSFSEPAQOAMRKALTLRYALLPHLYTLFHOAHVAGETVARPLFLE 1006
Db 476 SRHNHLGTIQEPWLPDVSVAEASRKALNRYTLPLVYTLFHEAHVSGLPVMEPLPFE 535
QY 1007 FPKDSSTWTVHQLLWGBALLITPVLOAKGAEVTVYFPLGTWYDLOTVPIEALGSLPPPP 1066
Db 536 FPDDAETVIDRQFLWGSALLVAPLEPGATSVKAYLPGGRWYDLYTGAGEA----- 587
QY 1067 AAPREPAIHSEGQWTLPAPLDTINVHLRAGVIIPLQGPGLTTTTSRQOPMALAVALTKG 1126
Db 588 -----SRGNVTLSAPLDKIPVHRGSGSIITQEPALTTTSDRONPFHLLVALDDN 638
QY 1127 GEARGELFWDDGESLEVLERGAYTOVIFLARNTIIVNELRVTV---SEGAGLQLOKVTVL 1183
Db 639 GTASGELYLDGSEIDT-QRGDYLLVQFSANNNTLTGTEV-VTGYVKNSTNLTLEKITL 696
QY 1184 GVATAPOQV 1192
Db 697 GVGNEPAAV 705

RESULT 5
US-09-819-247-2
; Sequence 2, Application US/09819247
; Patent No. US20010036635A1
; GENERAL INFORMATION:
; APPLICANT: Waldman, Scott A.
; APPLICANT: Park, Jason
; APPLICANT: Schulz, Stephanie
; TITLE OF INVENTION: Compositions And Methods For Identifying And Targeting Cancer Cell
; FILE REFERENCE: TJU2413
; CURRENT APPLICATION NUMBER: US/09/819,247
; CURRENT FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/192,229
; PRIOR FILING DATE: 2000-03-27
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1827
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-819-247-2

Query Match 28.8%; Score 1877.5; DB 9; Length 1827;
Best Local Similarity 41.7%; Pred. No. 4e-137;
Matches 382; Conservative 157; Mismatches 317; Indels 59; Gaps 19;
QY 470 PIETPRVHGRAPSLYVSFSEPFVHVRQDGRVLLNTTVAPLFFADQFLQLSTSLP 529
Db 1 PLDFPASSPAESDLYDLVNSNGPFGFEVIRKSTGDLFDFTTFFGLVFSQDLQLSTSLP 60
QY 530 SQVITGLAHLSPMLSTWTRITLWNRDLATPPG-ANLYGSHPPFVLALEDGSGAHGVFL 588
Db 61 SEYIIGLGEHAHKLFRDRTNETVTLWNRDVGPSYGDNNLYGSHPPFYMSLEDSGNAHGVFL 120

QY 1102 LQPGGLTTTSSRQOPMALAVALTGKGARGELFWDDGESLEVLERGAYTQVIFLARNNTI 1161
Db 799 IOEPDVTTTASRKNPLGLIIVGALGENTAKGDFWDDGETKDTIQNGNYLYITFSVSNNTL 858
QY 1162 VNELVRVTS---EGAGLOLQKVTVLGV--ATAPOQVLSNGVPV---SNFTYSPDTKVLDI 1213
Db 859 --DIVCTHSSYQEGTTLAFQTKILGLTDSVTEVRVAENNOPMNAHSNFTYDASNOVLII 916
QY 1214 C-VSLLMGEQFLVSW 1227
Db 917 ADLKLNLGRNFSVQW 931

RESULT 7
US-10-751-699-14
; Sequence 14, Application US/10751699
; Publication No. US20040142389A1
; GENERAL INFORMATION:
; APPLICANT: O'Mahony, Daniel J
; TITLE OF INVENTION: Antibodies To Peptides That Target GIT Receptors And Related Meth
; FILE REFERENCE: P26,477-B USA
; CURRENT APPLICATION NUMBER: US/10/751,699
; PRIOR FILING DATE: 2004-01-05
; PRIOR APPLICATION NUMBER: US 09/443,780
; PRIOR FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: US 60/109,036
; PRIOR FILING DATE: 1998-11-19
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 1827
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: hSI receptor
US-10-751-699-14

Query Match 28.8%; Score 1877.5; DB 16; Length 1827;
Best Local Similarity 41.7%; Pred. No. 4e-137;
Matches 382; Conservative 157; Mismatches 317; Indels 59; Gaps 19;

QY 340 ABETGAHGRRAVPTQCDVPNSRFDCAPKAITOEOCEARGCCYIPAKOGLQAQMG 399
Db 49 ATTRVTNPSDGSCKPNVLNDPVNVRINCIPQEPTEGICAGRGCCWRPNWDSL----- 102
QY 400 QNCEPPPPSPYKLENLSSENGYATLTR--TTPTFFPKDILTLRLDVMWMTENRLHFT 458
Db 103 IPWCFFVDNH-GYNVODMTTISIGVEAKLNRIPTSLFGNDINSVLFTTQNPNNRFRK 161
QY 459 IKDPANRRYEVEPLETPRVHS--RAPSPLYSVBFSEPPFGVIVHRQDGRVLLNTTVAPLF 516
Db 162 ITDPNNRRYEVEPHQYKEFTGPTVSDTLVDVKAQNPFSIQVIRKSNKTLFTDSIGPLV 221
QY 517 FADQFLQSTSLPSQVITGLAEHL-SPLMLSTSWTRITLWNRDLAP-TPCANLYGSHPPY 574
Db 222 YSDQYLQISARLPDSYIYGEGVQVRRHDLWSKWTPIFTRDQLPGDNNNNLYGHQTFP 281
QY 575 LALED-GGSAHGVELNNSAMDVLQPSALSWRSRGTGLDVYVIFLGPPEKSVQOQLDV 633
Db 282 MCIEDTSGKSGFGLNLSNAMEIFIQTPVTVRVGTGILDFFVILIGDTPQVQVQOQL 341
QY 634 VGYPFPMPYWGFLHLCRGGYSTATITQOVENMTTRAHPPLDVQMNLDLYMDSRDDFTFN 693
Db 342 VGLPAMPAYNLGFLQSRNYSKLVVVKEVVRNREAGIPFTQVTDIDYMEDKXDTYD 401
QY 694 KQGRFRPPAMQELHOGGRYRMVDPDAISSGPGAG--SYRYPDEGLRGVFTITNETGO- 750
Db 402 QVAFNGLPQFVDLHDHGQKYVILDPALISIGRRANGTYYATYERGNTQHVWINESDGT 461
QY 751 PLIGKVPGSTAFPPDFTNTALAWMEDVAEHDQVPPDGLWIDMNEPSNFIKRGSDGCP 810
Db 462 PIIGVEWPGLVTPFTNPNCDWMANECSIFHOEVQYDGLWIDMNEVSSFIQGTGKCN 521

QY 811 NNELENPYPVPGVGGTLOAATICASHQFISTHYNLNLNLYGLTEAISHRALVKA-RGT 869
Db 522 VNKLNPYPPTDILDKLMYSKTCICMDAVQNMCKQYDVHSLYGYSMATIEQAVQKVPFNK 581
QY 870 RPFVLSRSTFAGHGRYAGHTGCDWSSWEOLASSVPEILOFNLLGVPLVYADVCGFLGNT 929
Db 582 RSFILTRSTFAGSGRHAHAWLGDNTASWEQMEWSITGMLEFSLFGIPLVAGADICGFAET 641
QY 930 SEELCVRTQLCAGFYPPFMRNHNLSLPOEYPSFSEPA-----QQAMRKALTIRYAL 981
Db 642 TEELCRMWQLGAFYPPFSRNHS-----DGYEHQDPAFFGQNSLLVKSRYLIRYTL 695
QY 982 LPHLYTLFHOAHVAGETVARPLFLFPKDSSTWTVYDHLQHLGAEALLITPVLOAGKAEVTG 1041
Db 696 LPFLYTLFVKAHVFGETVARPVLHEFYEDTNSWIEDTEFLWGPALLITPVLKQGADTVSA 755
QY 1042 YPPLGTWYDLQVPIEALGSLPPPPAAPREPAIHSEGQWVTLPAPLDTINVHLRAGYIIP 1101
Db 756 YIPDAIWDYDES-----GAKRP-----WRKQVDMYLPADKIGLHLRGYIIP 798
QY 1102 LQPGGLTTTSSRQOPMALAVALTGKGARGELFWDDGESLEVLERGAYTQVIFLARNNTI 1161
Db 799 IOEPDVTTTASRKNPLGLIIVGALGENTAKGDFWDDGETKDTIQNGNYLYITFSVSNNTL 858
QY 1162 VNELVRVTS---EGAGLOLQKVTVLGV--ATAPOQVLSNGVPV---SNFTYSPDTKVLDI 1213
Db 859 --DIVCTHSSYQEGTTLAFQTKILGLTDSVTEVRVAENNOPMNAHSNFTYDASNOVLII 916
QY 1214 C-VSLLMGEQFLVSW 1227
Db 917 ADLKLNLGRNFSVQW 931

RESULT 8
US-10-600-862A-9
; Sequence 9, Application US/10600862A
; Publication No. US20050026823A1
; GENERAL INFORMATION:
; APPLICANT: Zankel et al.
; TITLE OF INVENTION: USE OF THE CHAPERONE RECEPTOR-ASSOCIATED PROTEIN (RAP) FOR THE
; FILE REFERENCE: 30610/39383
; CURRENT APPLICATION NUMBER: US/10/600,862A
; CURRENT FILING DATE: 2003-06-20
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 972
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: RAP-1DU fusion sequence
US-10-600-862A-9

Query Match 27.3%; Score 1783; DB 17; Length 972;
Best Local Similarity 99.4%; Pred. No. 3.6e-130;
Matches 346; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MRGPSGALWLLALRTVLGYSREKNQPKSPKRESGEFFRMEKLNQLWEKAQRLHLPV 60
Db 1 MRGPSGALWLLALRTVLGYSREKNQPKSPKRESGEFFRMEKLNQLWEKAQRLHLPV 60
QY 61 RLAEHLADLKIOERDELAWKKLKLDDGDEGEKEARLIRNLNVLAKYGLDGKQDARQVT 120
Db 61 RLAEHLADLKIOERDELAWKKLKLDDGDEGEKEARLIRNLNVLAKYGLDGKQDARQVT 120
QY 121 SNSLSGTQEDGLDDPRLEKLMHAKTSKFGSEELDKLWREPLHHKKEVHENVLLETIS 180
Db 121 SNSLSGTQEDGLDDPRLEKLMHAKTSKFGSEELDKLWREPLHHKKEVHENVLLETIS 180
QY 181 RTEETHENVISPSDLSDKGSLVSHRHTLEKLSINOGDLRLRRVSHOQYSTAEFEE 240

Db 181 RTEEIHENVISPSDLSDIKGSLVLSRHTLKEKLSINQGLDRLRRVSHQGYSTAEFEE 240
Qy 241 PRVIDLWDLAQSANLTKLEAFREELKHFEAKIEKHNYQKOLEIAHEKLRHAESVGDG 300
Db 241 PRVIDLWDLAQSANLTKLEAFREELKHFEAKIEKHNYQKOLEIAHEKLRHAESVGDG 300
Qy 301 ERVSRREKHALLGRTKELGYTVKXHLQDLSGRISRARAETGAHP 348
Db 301 ERVSRREKHALLGRTKELGYTVKXHLQDLSGRISRARAETGAHP 348

RESULT 9
US-10-812-849-9
; Sequence 9, Application US/10812849
; Publication No. US20050042227A1
; GENERAL INFORMATION:
; APPLICANT: Zankel et al.
; TITLE OF INVENTION: MEGALIN-BASED DELIVERY OF THERAPEUTIC COMPOUNDS TO THE BRAIN AND OTHER TISSUES
; FILE REFERENCE: 30610/40037
; CURRENT APPLICATION NUMBER: US/10/812,849
; PRIOR FILING DATE: 2004-03-30
; PRIOR APPLICATION NUMBER: US 10/600,862
; PRIOR FILING DATE: 2003-06-20
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: RAP-1DU fusion sequence
US-10-812-849-9

Query Match 27.3%; Score 1783; DB 17; Length 972;
Best Local Similarity 99.4%; Pred. No. 3.6e-130;
Matches 346; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
Qy 1 MRGPSALWLLALRTVLGYSREKQPKSPKRESGEFRMEKLNQWKAQRLHLPV 60
Db 1 MRGPSALWLLALRTVLGYSREKQPKSPKRESGEFRMEKLNQWKAQRLHLPV 60
Qy 61 RLAEHLADLKIQRDELAWKLLDGLDDEGEKEARLIRNLNVLAKYGLDGKKDQV 120
Db 61 RLAEHLADLKIQRDELAWKLLDGLDDEGEKEARLIRNLNVLAKYGLDGKKDQV 120
Qy 121 SNSLSGTQEDGLDPPLEKLWHKAKTSGKPSGELDKLWREFLHHKEKVHEYNVLTLS 180
Db 121 SNSLSGTQEDGLDPPLEKLWHKAKTSGKPSGELDKLWREFLHHKEKVHEYNVLTLS 180
Qy 181 RTEEIHENVISPSDLSDIKGSLVLSRHTLKEKLSINQGLDRLRRVSHQGYSTAEFEE 240
Db 181 RTEEIHENVISPSDLSDIKGSLVLSRHTLKEKLSINQGLDRLRRVSHQGYSTAEFEE 240
Qy 241 PRVIDLWDLAQSANLTKLEAFREELKHFEAKIEKHNYQKOLEIAHEKLRHAESVGDG 300
Db 241 PRVIDLWDLAQSANLTKLEAFREELKHFEAKIEKHNYQKOLEIAHEKLRHAESVGDG 300
Qy 301 ERVSRREKHALLGRTKELGYTVKXHLQDLSGRISRARAETGAHP 348
Db 301 ERVSRREKHALLGRTKELGYTVKXHLQDLSGRISRARAETGAHP 348

RESULT 10
US-10-600-862a-11
; Sequence 11, Application US/10600862a
; Publication No. US20050026823A1
; GENERAL INFORMATION:
; APPLICANT: Zankel et al.
; TITLE OF INVENTION: USE OF THE CHAPERONE RECEPTOR-ASSOCIATED PROTEIN (RAP) FOR THE DELIVERY OF THERAPEUTIC COMPOUNDS TO THE BRAIN AND OTHER TISSUES
; FILE REFERENCE: 30610/39383
; CURRENT APPLICATION NUMBER: US/10/600,862a

; CURRENT FILING DATE: 2003-06-20
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: RAP-GDNF fusion sequence
US-10-600-862a-11

Query Match 26.1%; Score 1702; DB 17; Length 463;
Best Local Similarity 90.6%; Pred. No. 2.4e-124;
Matches 338; Conservative 4; Mismatches 19; Indels 12; Gaps 2;
Qy 19 GSYSREKQPKSPKRESGEFRMEKLNQWKAQRLHLPVRLAEHLADLKIQRDELA 78
Db 3 GSYSREKQPKSPKRESGEFRMEKLNQWKAQRLHLPVRLAEHLADLKIQRDELA 62
Qy 79 WKLLKLDGLDEGEKEARLIRNLNVLAKYGLDGKKDQVTSNSLSGTQEDGLDPPLE 138
Db 63 WKLLKLDGLDEGEKEARLIRNLNVLAKYGLDGKKDQVTSNSLSGTQEDGLDPPLE 122
Qy 139 KLWHKAKTSGKPSGELDKLWREFLHHKEKVHEYNVLTLSRTEEIHENVISPSLSDI 198
Db 123 KLWHKAKTSGKPSGELDKLWREFLHHKEKVHEYNVLTLSRTEEIHENVISPSLSDI 182
Qy 199 KGSVLHSRHTLKEKLSINQGLDRLRRVSHQGYSTAEFEPVRVIDLWDLAQSANLTK 258
Db 183 KGSVLHSRHTLKEKLSINQGLDRLRRVSHQGYSTAEFEPVRVIDLWDLAQSANLTK 242
Qy 259 ELEAFREELKHFEAKIEKHNYQKOLEIAHEKLRHAESVGDGERVSRREKHALLEGRTK 318
Db 243 ELEAFREELKHFEAKIEKHNYQKOLEIAHEKLRHAESVGDGERVSRREKHALLEGRTK 302
Qy 319 ELGYTVKXHLQDLSGRISRARAETGAH-----PGRPRAVPTQCDVPPNSRFDCA 371
Db 303 ELGYTVKXHLQDLSGRISRARAETGSPDKOMAVLPRERNRQAAAAANPENS 357
Qy 372 KAITOQCEARGC 384
Db 358 KGRGQRGNRC 370

RESULT 11
US-10-812-849-11
; Sequence 11, Application US/10812849
; Publication No. US20050042227A1
; GENERAL INFORMATION:
; APPLICANT: Zankel et al.
; TITLE OF INVENTION: MEGALIN-BASED DELIVERY OF THERAPEUTIC COMPOUNDS TO THE BRAIN AND OTHER TISSUES
; FILE REFERENCE: 30610/40037
; CURRENT APPLICATION NUMBER: US/10/812,849
; CURRENT FILING DATE: 2004-03-30
; PRIOR APPLICATION NUMBER: US 10/600,862
; PRIOR FILING DATE: 2003-06-20
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: RAP-GDNF fusion sequence
US-10-812-849-11

Query Match 26.1%; Score 1702; DB 17; Length 463;
Best Local Similarity 90.6%; Pred. No. 2.4e-124;
Matches 338; Conservative 4; Mismatches 19; Indels 12; Gaps 2;
Qy 19 GSYSREKQPKSPKRESGEFRMEKLNQWKAQRLHLPVRLAEHLADLKIQRDELA 78

Db 3 GSYSRKNQPKSPKRESGEFRMEKLNQWKAQRLHLPVRLAELHADLKIQRDELA 62

Qy 79 WKKLKLDGLDDEGEKEARLIRNLNVLAKYGLDGKDKARQVTSNSLSGTQEDGLDPRLE 138

Db 63 WKKLKLDGLDDEGEKEARLIRNLNVLAKYGLDGKDKARQVTSNSLSGTQEDGLDPRLE 122

Qy 139 KLWHKAKTSKGFSGEELDKLWREFLHHKVKVHEYNVLLTSLRTEIHNENVISPSDLSI 198

Db 123 KLWHKAKTSKGFSGEELDKLWREFLHHKVKVHEYNVLLTSLRTEIHNENVISPSDLSI 182

Qy 199 KGSVLHSRHTELKEKLSRNOGLDRLRRVSHQSYSTEAFEEPRVLDLWDLAQSANLTK 258

Db 183 KGSVLHSRHTELKEKLSRNOGLDRLRRVSHQSYSTEAFEEPRVLDLWDLAQSANLTK 242

Qy 259 ELEAFREELKHFAKIEKNHNYQKLEIAHEKLRHAESVGDGERSRSREKHALLEGRTK 318

Db 243 ELEAFREELKHFAKIEKNHNYQKLEIAHEKLRHAESVGDGERSRSREKHALLEGRTK 302

Qy 319 ELGYTVKKHLQDLSGRISRARAETGAH-----PGRPRAVPTQCDVPPNSRFDPCAPD 371

Db 303 ELGYTVKKHLQDLSGRISRARAETGSPDKQMAVLPRERNRQAAAAANPENS-----G 357

Qy 372 KAITQEQCEARGC 384

Db 358 KGRRGQRGNRG 370

RESULT 12

US-10-600-862A-21

Query Match 25.5%; Score 1660.5; DB 17; Length 357;

Best Local Similarity 95.1%; Pred. No. 2.8e-121;

Matches 327; Conservative 2; Mismatches 10; Indels 5; Gaps 1;

GENERAL INFORMATION:

APPLICANT: Zankel et al.

TITLE OF INVENTION: USE OF THE CHAPERONE RECEPTOR-ASSOCIATED PROTEIN (RAP) FOR THE DELIVERY OF THERAPEUTIC COMPOUNDS TO THE BRAIN AND OTHER TISSUES

FILE REFERENCE: 30610/39383

CURRENT APPLICATION NUMBER: US/10/600,862A

CURRENT FILING DATE: 2003-06-20

NUMBER OF SEQ ID NOS: 28

SOFTWARE: PatentIn version 3.2

SEQ ID NO 21

TYPE: PRT

ORGANISM: Homo sapiens

US-10-600-862A-21

Qy 1 MRGPGALWLLAL-----RTVLGSGYSREKNQPKSPKRESGEFRMEKLNQWKAQRL 55

Db 10 LRGLPALLLLLFLGWPWPAASHGGKYSREKNQPKSPKRESGEFRMEKLNQWKAQRL 69

Qy 56 HLPVRLAELHADLKIQRDELAWKKLKLDGLDDEGEKEARLIRNLNVLAKYGLDGKDK 115

Db 70 HLPVRLAELHADLKIQRDELAWKKLKLDGLDDEGEKEARLIRNLNVLAKYGLDGKDK 129

Qy 116 ARQVTSNSLSGTQEDGLDPRLEKLWHKAKTSKGFSGEELDKLWREFLHHKVKVHEYNVL 175

Db 130 ARQVTSNSLSGTQEDGLDPRLEKLWHKAKTSKGFSGEELDKLWREFLHHKVKVHEYNVL 189

Qy 176 LETLSRTEIHNENVISPSDLSIKGSLVLSRHTELKEKLSRNOGLDRLRRVSHQSYSTE 235

Db 190 LETLSRTEIHNENVISPSDLSIKGSLVLSRHTELKEKLSRNOGLDRLRRVSHQSYSTE 249

Qy 236 ABEFEPRIIDLWDLAQSANLTKLEAFREELKHFAKIEKNHNYQKLEIAHEKLRHAE 295

Db 250 ABEFEPRIIDLWDLAQSANLTKLEAFREELKHFAKIEKNHNYQKLEIAHEKLRHAE 309

Qy 296 SVGDGERSRSREKHALLEGRTKELGYTVKKHLQDLSGRISRAR 339

Db 310 SVGDGERSRSREKHALLEGRTKELGYTVKKHLQDLSGRISRAR 353

RESULT 13

US-10-868-577A-45

Query Match 25.5%; Score 1660.5; DB 17; Length 357;

Best Local Similarity 95.1%; Pred. No. 2.8e-121;

Matches 327; Conservative 2; Mismatches 10; Indels 5; Gaps 1;

GENERAL INFORMATION:

APPLICANT: Alitalo et al.

TITLE OF INVENTION: HEPARIN BINDING VEGFR-3 LIGANDS

FILE REFERENCE: 28967/39359A

CURRENT APPLICATION NUMBER: US/10/868,577A

CURRENT FILING DATE: 2004-06-14

PRIOR APPLICATION NUMBER: US 60/478,390

PRIOR FILING DATE: 2003-06-12

PRIOR APPLICATION NUMBER: US 10/669,176

PRIOR FILING DATE: 2003-09-23

NUMBER OF SEQ ID NOS: 69

SOFTWARE: PatentIn version 3.2

SEQ ID NO 45

TYPE: PRT

ORGANISM: Homo sapiens

US-10-868-577A-45

Qy 1 MRGPGALWLLAL-----RTVLGSGYSREKNQPKSPKRESGEFRMEKLNQWKAQRL 55

Db 10 LRGLPALLLLLFLGWPWPAASHGGKYSREKNQPKSPKRESGEFRMEKLNQWKAQRL 69

Qy 56 HLPVRLAELHADLKIQRDELAWKKLKLDGLDDEGEKEARLIRNLNVLAKYGLDGKDK 115

Db 70 HLPVRLAELHADLKIQRDELAWKKLKLDGLDDEGEKEARLIRNLNVLAKYGLDGKDK 129

Qy 116 ARQVTSNSLSGTQEDGLDPRLEKLWHKAKTSKGFSGEELDKLWREFLHHKVKVHEYNVL 175

Db 130 ARQVTSNSLSGTQEDGLDPRLEKLWHKAKTSKGFSGEELDKLWREFLHHKVKVHEYNVL 189

Qy 176 LETLSRTEIHNENVISPSDLSIKGSLVLSRHTELKEKLSRNOGLDRLRRVSHQSYSTE 235

Db 190 LETLSRTEIHNENVISPSDLSIKGSLVLSRHTELKEKLSRNOGLDRLRRVSHQSYSTE 249

Qy 236 ABEFEPRIIDLWDLAQSANLTKLEAFREELKHFAKIEKNHNYQKLEIAHEKLRHAE 295

Db 250 ABEFEPRIIDLWDLAQSANLTKLEAFREELKHFAKIEKNHNYQKLEIAHEKLRHAE 309

Qy 296 SVGDGERSRSREKHALLEGRTKELGYTVKKHLQDLSGRISRAR 339

Db 310 SVGDGERSRSREKHALLEGRTKELGYTVKKHLQDLSGRISRAR 353

RESULT 14

US-10-868-577A-53

Query Match 25.5%; Score 1660.5; DB 17; Length 357;

Best Local Similarity 95.1%; Pred. No. 2.8e-121;

Matches 327; Conservative 2; Mismatches 10; Indels 5; Gaps 1;

GENERAL INFORMATION:

APPLICANT: Alitalo et al.

TITLE OF INVENTION: HEPARIN BINDING VEGFR-3 LIGANDS

FILE REFERENCE: 28967/39359A

CURRENT APPLICATION NUMBER: US/10/868,577A

CURRENT FILING DATE: 2004-06-14

PRIOR APPLICATION NUMBER: US 60/478,390

PRIOR FILING DATE: 2003-06-12

PRIOR APPLICATION NUMBER: US 10/669,176

PRIOR FILING DATE: 2003-09-23

NUMBER OF SEQ ID NOS: 69

SOFTWARE: PatentIn version 3.2

SEQ ID NO 53

TYPE: PRT

ORGANISM: Homo sapiens

US-10-868-577A-53

Qy 1 MRGPGALWLLAL-----RTVLGSGYSREKNQPKSPKRESGEFRMEKLNQWKAQRL 55

Db 10 LRGLPALLLLLFLGWPWPAASHGGKYSREKNQPKSPKRESGEFRMEKLNQWKAQRL 69

Qy 56 HLPVRLAELHADLKIQRDELAWKKLKLDGLDDEGEKEARLIRNLNVLAKYGLDGKDK 115

Db 70 HLPVRLAELHADLKIQRDELAWKKLKLDGLDDEGEKEARLIRNLNVLAKYGLDGKDK 129

Qy 116 ARQVTSNSLSGTQEDGLDPRLEKLWHKAKTSKGFSGEELDKLWREFLHHKVKVHEYNVL 175

Db 130 ARQVTSNSLSGTQEDGLDPRLEKLWHKAKTSKGFSGEELDKLWREFLHHKVKVHEYNVL 189

Qy 176 LETLSRTEIHNENVISPSDLSIKGSLVLSRHTELKEKLSRNOGLDRLRRVSHQSYSTE 235

Db 190 LETLSRTEIHNENVISPSDLSIKGSLVLSRHTELKEKLSRNOGLDRLRRVSHQSYSTE 249

Qy 236 ABEFEPRIIDLWDLAQSANLTKLEAFREELKHFAKIEKNHNYQKLEIAHEKLRHAE 295

Db 250 ABEFEPRIIDLWDLAQSANLTKLEAFREELKHFAKIEKNHNYQKLEIAHEKLRHAE 309

Qy 296 SVGDGERSRSREKHALLEGRTKELGYTVKKHLQDLSGRISRAR 339

Db 310 SVGDGERSRSREKHALLEGRTKELGYTVKKHLQDLSGRISRAR 353

US-10-868-577A-53

Query Match 25.5%; Score 1660.5; DB 17; Length 357;
Best Local Similarity 95.1%; Pred. No. 2.8e-121;
Matches 327; Conservative 2; Mismatches 10; Indels 5; Gaps 1;
QY 1 MRGPGALWLLAL-----RTVLGSYSREKNQPKSPKRESGEFRMEKLNQLWEKAQRL 55
DB 10 LRGLPALLLLFLGWPWPAASHGCKYSREKNQPKSPKRESGEFRMEKLNQLWEKAQRL 69
QY 56 HLPVRLAEHLADLKIQRDELAWKKLKDGLDDEGEKEARLIRNLNVLAKYGLDGKKD 115
DB 70 HLPVRLAEHLADLKIQRDELAWKKLKDGLDDEGEKEARLIRNLNVLAKYGLDGKKD 129
QY 116 ARQVTSNSLSGTQEDGLDPRLEKLMHKAKTSGKFSGEELDKLWREFLHHKEKVHEYNYL 175
DB 130 ARQVTSNSLSGTQEDGLDPRLEKLMHKAKTSGKFSGEELDKLWREFLHHKEKVHEYNYL 189
QY 176 LETLSRTEIHNHENVISPSDLSDIKGSVLHSRHTLKEKLRINSNOGLDRLRRVSHQGYSYE 235
DB 190 LETLSRTEIHNHENVISPSDLSDIKGSVLHSRHTLKEKLRINSNOGLDRLRRVSHQGYSYE 249
QY 236 AFEFEPVIDLWDLAQSANLTDKELEAFREELKHFEAKIEKKNHYOKOLEIAHEKLRHAE 295
DB 250 AFEFEPVIDLWDLAQSANLTDKELEAFREELKHFEAKIEKKNHYOKOLEIAHEKLRHAE 309
QY 296 SVGDGERVSRSEKHALLEGRTKELGYTVKKHLQDLSGRISRAR 339
DB 310 SVGDGERVSRSEKHALLEGRTKELGYTVKKHLQDLSGRISRAR 353

RESULT 15

US-10-812-849-21
; Sequence 21, Application US/10812849
; Publication No. US20050042227A1
; GENERAL INFORMATION:
; APPLICANT: Zankel et al.
; TITLE OF INVENTION: MEGALIN-BASED DELIVERY OF THERAPEUTIC COMPOUNDS TO THE BRAIN AND
; FILE REFERENCE: 30610/40037
; CURRENT APPLICATION NUMBER: US/10/812,849
; CURRENT FILING DATE: 2004-03-30
; PRIOR APPLICATION NUMBER: US 10/600,862
; PRIOR FILING DATE: 2003-06-20
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 21
; LENGTH: 357
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-812-849-21

Query Match 25.5%; Score 1660.5; DB 17; Length 357;
Best Local Similarity 95.1%; Pred. No. 2.8e-121;
Matches 327; Conservative 2; Mismatches 10; Indels 5; Gaps 1;
QY 1 MRGPGALWLLAL-----RTVLGSYSREKNQPKSPKRESGEFRMEKLNQLWEKAQRL 55
DB 10 LRGLPALLLLFLGWPWPAASHGCKYSREKNQPKSPKRESGEFRMEKLNQLWEKAQRL 69
QY 56 HLPVRLAEHLADLKIQRDELAWKKLKDGLDDEGEKEARLIRNLNVLAKYGLDGKKD 115
DB 70 HLPVRLAEHLADLKIQRDELAWKKLKDGLDDEGEKEARLIRNLNVLAKYGLDGKKD 129
QY 116 ARQVTSNSLSGTQEDGLDPRLEKLMHKAKTSGKFSGEELDKLWREFLHHKEKVHEYNYL 175
DB 130 ARQVTSNSLSGTQEDGLDPRLEKLMHKAKTSGKFSGEELDKLWREFLHHKEKVHEYNYL 189
QY 176 LETLSRTEIHNHENVISPSDLSDIKGSVLHSRHTLKEKLRINSNOGLDRLRRVSHQGYSYE 235
DB 190 LETLSRTEIHNHENVISPSDLSDIKGSVLHSRHTLKEKLRINSNOGLDRLRRVSHQGYSYE 249
QY 236 AFEFEPVIDLWDLAQSANLTDKELEAFREELKHFEAKIEKKNHYOKOLEIAHEKLRHAE 295

DB 250 AFEFEPVIDLWDLAQSANLTDKELEAFREELKHFEAKIEKKNHYOKOLEIAHEKLRHAE 309
QY 296 SVGDGERVSRSEKHALLEGRTKELGYTVKKHLQDLSGRISRAR 339
DB 310 SVGDGERVSRSEKHALLEGRTKELGYTVKKHLQDLSGRISRAR 353

Search completed: April 5, 2005, 01:43:08
Job time : 1604 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2005, 22:21:01 ; Search time 52 Seconds
(without alignments)
2272.196 Million cell updates/sec

Title: US-10-600-862A-7

Perfect score: 6522
Sequence: 1 MRGPGALWLLALRLVLGS.....KVLDCVSLLMGEQFLVSNW 1228

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

PIR 79:.*
1: PIR1:.*
2: PIR2:.*
3: PIR3:.*
4: PIR4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4731.5	72.5	952	1 A32609	alpha-glucosidase
2	1877.5	28.8	1827	1 U0HU	sucrose alpha-gluc
3	1869.5	28.7	1827	1 A23945	sucrose alpha-gluc
4	1802.5	27.6	1841	2 T10799	sucrose alpha-gluc
5	1660.5	25.5	357	2 A39875	alpha-2-macroglobu
6	1522.5	23.3	902	2 T48531	alpha-glucosidase
7	1494.5	22.9	903	2 T09143	alpha-glucosidase
8	1458.5	22.4	913	2 JC5463	alpha-glucosidase
9	1449.5	22.2	915	2 H96709	hypothetical prote
10	1416	21.7	1743	2 T15893	hypothetical prote
11	1403.5	21.5	877	2 S65057	alpha-glucosidase
12	1380	21.2	993	2 T38598	probable family 31
13	1361.5	20.9	995	2 T50267	probable family 31
14	1330.5	20.4	864	2 JC4624	alpha-glucosidase
15	1327.5	20.4	360	2 JX0281	alpha-2-macroglobu
16	1325.5	20.3	357	2 A46646	Heymann nephritis
17	1306	20.0	855	2 T47534	hypothetical prote
18	1263	19.4	919	2 T16693	hypothetical prote
19	1262	19.3	985	2 JC4217	alpha-glucosidase
20	1232.5	18.9	958	1 JN0102	glucan 1,4-alpha-g
21	1215	18.6	856	2 T22575	hypothetical prote
22	1117	17.1	719	2 JC1200	alpha-glucosidase
23	1090	16.7	1070	2 S19686	alpha-glucosidase
24	895	13.7	779	2 AE2402	alpha-glucosidase
25	891.5	13.7	919	2 T07391	probable alpha-glu
26	891.5	13.7	941	2 T32449	hypothetical prote
27	882	13.5	763	2 AH1097	alpha-glucosidase
28	879	13.5	763	2 AG1460	alpha-glucosidase
29	831	12.7	954	2 S46105	glucan 1,4-alpha-g

RESULT 1

A32609

Alpha-glucosidase (EC 3.2.1.20) precursor, lysosomal - human
N:Alternate names: acid alpha-glucosidase; acid maltase; maltase-glucoamylase
C:Species: Homo sapiens (man)
C:Date: 21-May-1990 #sequence revision 05-Apr-1995 #text change 18-Jun-1999
C:Accession: A40577; A32609; A35698; S00831; S18847; I52309; S63526
R:Martiniuk, F.; Bodkin, M.; Tzall, S.; Hirschhorn, R.
DNA Cell Biol. 10, 283-292, 1991

A:Title: Isolation and partial characterization of the structural gene for human acid al
A:Reference number: A40577; MUID:91229698; PMID:1674202
A:Accession: A40577

A:Status: not compared with conceptual translation
A:Molecule type: DNA

A:Residues: 1-952 <MA2>
A>Note: order of exons in Figure 2 is incorrect

R:Martiniuk, F.; Mehler, M.; Tzall, S.; Meredith, G.; Hirschhorn, R.
DNA Cell Biol. 9, 85-94, 1990

A:Title: Sequence of the cDNA and 5'-flanking region for human acid alpha-glucosidase, de
s with previous cDNA and amino acid sequences.

A:Reference number: A32609; MUID:90262851; PMID:2111708
A:Accession: A32609

A:Molecule type: mRNA
A:Residues: 1-952 <MA2>

A:Cross-references: GB:M34424; NID:g182907; PIDN:AAA52506.1; PID:g182908
R:Martiniuk, F.; Bodkin, M.; Tzall, S.; Hirschhorn, R.

Am. J. Hum. Genet. 47, 440-445, 1990
A:Title: Identification of the base-pair substitution responsible for a human acid alpha

A:Reference number: A35698; MUID:90365036; PMID:2203258
A:Accession: A35698

A:Contents: partial sequence of GAA 2 allelic form
A:Status: not compared with conceptual translation

A:Molecule type: mRNA
A:Residues: 1-90, 'N', 92-129 <MA3>

R:Hoefsloot, L.H.; Hoogeveen-Westerveld, M.; Kroos, M.A.; van Beeumen, J.; Reuser, A.J.J.
EMBO J. 7, 1697-1704, 1988

A:Title: Primary structure and processing of lysosomal alpha-glucosidase; homology with t
A:Reference number: S00831; MUID:89005058; PMID:3049072
A:Accession: S00831

A:Molecule type: mRNA
A:Residues: 1-136, 'VLL', 130-198, 'R', 200-222, 'H', 224-371, 'L', 373-401, 'R', 403-412, 'TSRSTRM

A:Cross-references: EMBL:X00839
A:Accession: S18847

A:Molecule type: protein
A:Residues: 70-89; 123-126, 'VLL', 130-145; 204-215; 230-249; 332-345; 349-401, 'R', 403-4

R:Lin, C.Y.; Shieh, J.J.
Biochem. Biophys. Res. Commun. 208, 886-893, 1995

A:Title: Identification of a de novo point mutation resulting in infantile form of Pompe'
A:Reference number: I52309; MUID:95209708; PMID:7695647
A:Accession: I52309

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA
A:Residues: 631-644, 'H', 646-680 <RES>
A:Cross-references: GB:S76893; NID:G912786; PIDN:AAB33842.1; PID:G912787
A>Note: homozygous mutation of Asp-645 to His causes an infantile form of Pompe's disease
R:Fuller, M.; van der Ploeg, A.; Reuser, A.J.J.; Anson, D.S.; Hopwood, J.J.
Eur. J. Biochem. 234, 903-909, 1995
A:Title: Isolation and characterization of a recombinant, precursor form of lysosomal ac
A:Reference number: S63526; MUID:96163476; PMID:8575451
A:Accession: S63526
A:Molecule type: protein
A:Residues: 68-74, 'X', 76-79, 'X', 81 <FUL>
C:Comment: In common with other lysosomal enzymes, posttranslational processing includes
processing at both amino and carboxyl ends.
C:Genetics:
A:Gene: GDB:GAA
A:Cross-references: GDB:115965; OMIM:232300
A:Map position: 17q23-17q23
A:Introns: 182/3; 231/2; 286/3; 319/1; 359/1; 398/3; 442/3; 479/3; 517/3; 546/1; 585/2;
A>Note: GAA 1, GAA 2, and GAA 4 are common alleles in the normal population
C:Function:
A:Description: also has EC 3.1.2.3 activity; hydrolyzes alpha 1,4 and alpha 1,6 glycosid
A>Note: deficiency causes glycogen storage disease type II (Pompe's disease)
C:Superfamily: lysosomal alpha-glucosidase; sucrase/isomaltase homology; trefoil homology
C:Keywords: glycoprotein; glycosidase; hydrolyase; lysosomal storage disease; lysosome; H
F:82-130/Domain: trefoil homology <TRF>
F:208-867/Domain: sucrase/isomaltase homology <SIM>
F:59-70/Cleavage site: Gln-Ala (unidentified proteinase) #status experimental
F:82-109,92-108,103-127/Dialdide bonds: #status predicted
F:122-123/Cleavage site: Met-Gly (unidentified proteinase) #status experimental
F:140,233/Binding site: carbohydate (Asn) (covalent) #status experimental
F:203-204/Cleavage site: Arg-Ala (unidentified proteinase) #status experimental
F:390,470,492,652,925/Binding site: carbohydate (Asn) (covalent) #status predicted
F:518/Active site: Asp #status predicted

Query Match 72.5%; Score 4731.5; DB 1; Length 952;
Best Local Similarity 96.9%; Pred. No. 2.8e-280;
Matches 888; Conservative 3; Mismatches 20; Indels 5; Gaps 1;

QY 313 LEGRTKELGYTVKHLQDLISGRISRAAEATGAHPGRPRVPTQCDVPNPFDFCAPDK 372
DB LSGSSPVLSETHPAHQGASRPGPR-----DAQNHGPRPRVPTQCDVPNPFDFCAPDK 96
QY 373 AITQEQCEARGCCYIPAKOGLQAQMGQPCWCPFPSPYKYKLENLSSSEMGYTATLTRTT 432
DB AITQEQCEARGCCYIPAKOGLQAQMGQPCWCPFPSPYKYKLENLSSSEMGYTATLTRTT 156
QY 433 PTFPPKDLTLRLDVMMETENRLHFTIKDPANRRYEVPLETRVHSHRAPSPLYSEFSEE 492
DB PTFPPKDLTLRLDVMMETENRLHFTIKDPANRRYEVPLETRVHSHRAPSPLYSEFSEE 216
QY 493 PFGVIVHRQLDGRVLNTTVAFLFFADQFLQLSTSLPSQYITGLAEHLSPMLSTSWTRI 552
DB PFGVIVHRQLDGRVLNTTVAFLFFADQFLQLSTSLPSQYITGLAEHLSPMLSTSWTRI 276
QY 553 TLNWRDLAPTPGANLVGSHPPFYALBEDGSGAHGVFLNNSNAMDVLQPSPALSWRSTGSI 612
DB TLNWRDLAPTPGANLVGSHPPFYALBEDGSGAHGVFLNNSNAMDVLQPSPALSWRSTGSI 336
QY 613 LDVYIFLGPBPKSVQOQLDVGVYRPMPPYWGIGHLCHWGYSSTAITRQVVENMTRAHP 672
DB LDVYIFLGPBPKSVQOQLDVGVYRPMPPYWGIGHLCHWGYSSTAITRQVVENMTRAHP 396
QY 673 PLDVQWNLDDYMSRRDFTFNKDGFRDFFPMVQELHQGGRYMMIVDPAISSSGPAGSYR 732
DB PLDVQWNLDDYMSRRDFTFNKDGFRDFFPMVQELHQGGRYMMIVDPAISSSGPAGSYR 456
QY 733 PYDEGLRGVFTINETGQPLIGKWPFGSTAFDFTNPTALAWEDVVAEFHQVDFDGLW 792
DB PYDEGLRGVFTINETGQPLIGKWPFGSTAFDFTNPTALAWEDVVAEFHQVDFDGLW 516
QY 793 IDNNEPSNFRGSDGCPNNELENPPYVGVGTTGTLQAATICASSHQFLSTHYNLNLGY 852
DB IDNNEPSNFRGSDGCPNNELENPPYVGVGTTGTLQAATICASSHQFLSTHYNLNLGY 576

QY 853 LTEAIAASHRALVKARGTRFPVIRSTFAGHGRYAGHWTDVWSSWEQLASSVPEILQFNL 912
DB LTEAIAASHRALVKARGTRFPVIRSTFAGHGRYAGHWTDVWSSWEQLASSVPEILQFNL 636
QY 913 LGVPLVGADVCGFLNGTSEELCVRTQTLGAFYPPFMRNHNLSLSLQEPYSFSEPAQAQNR 972
DB LGVPLVGADVCGFLNGTSEELCVRTQTLGAFYPPFMRNHNLSLSLQEPYSFSEPAQAQNR 696
QY 973 KALTURYALLPHLYTLFPHQAHVAGTVARPLFLBPPKDSSTWTVDHQLLWGSALLITPVL 1032
DB KALTURYALLPHLYTLFPHQAHVAGTVARPLFLBPPKDSSTWTVDHQLLWGSALLITPVL 756
QY 1033 QAGKAEVTGYFPLGTWYDLQTVPIEALGSLPPPPAAPREPAIHSEGGQWVTLPAPLDTINV 1092
DB QAGKAEVTGYFPLGTWYDLQTVPIEALGSLPPPPAAPREPAIHSEGGQWVTLPAPLDTINV 816
QY 1093 HLRAGYIIPLOQPGILTTTSSROOPMALAVALTKGGEARGELFWDDGESLEVLERGAYTQV 1152
DB HLRAGYIIPLOQPGILTTTSSROOPMALAVALTKGGEARGELFWDDGESLEVLERGAYTQV 876
QY 1153 IFLARNNTIIVNELVRVTSEAGLQLOKQTVLGAVATAPOQVLSNGVPSNFTYSPDTKVLID 1212
DB IFLARNNTIIVNELVRVTSEAGLQLOKQTVLGAVATAPOQVLSNGVPSNFTYSPDTKVLID 936
QY 1213 ICVSLLMGEQFLVSWC 1228
DB ICVSLLMGEQFLVSWC 952

RESULT 2
UUHU
sucrose alpha-glucosidase (EC 3.2.1.48) / oligo-1,6-glucosidase (EC 3.2.1.10) [validated]
N:Alternate names: limit dextrinase; small intestinal sucrase/isomaltase (SI)
C:Species: Homo sapiens (man)
C:Date: 19-Nov-1988 #sequence revision 24-May-1996 #text_change 09-Jul-2004
C:Accession: S36082; A27326; S24329; A61136
R:Lacasa, M.
submitted to the EMBL Data Library, December 1991
A:Reference number: S36082
A:Accession: S36082
A:Molecule type: mRNA
A:Residues: 1-1827 <LAC>
A:Cross-references: UNIPROT:P14410; EMBL:X63597; NID:G36644; PIDN:CAA45140.1; PID:G36645
R:Green, F.; Edwards, Y.; Hauri, H.P.; Povey, S.; Ho, M.W.; Pinto, M.; Swallow, D.
Gene 57, 101-110, 1987
A:Title: Isolation of a cDNA probe for a human jejunal brush-border hydrolase, sucrase-is
A:Reference number: A27326; MUID:88112852; PMID:2962903
A:Accession: A27326
A:Molecule type: mRNA
A:Residues: 1-661, 'X', 663-678 <GRE>
A:Cross-references: GB:M22616
R:Chantret, I.; Lacasa, M.; Chevallier, G.; Ruf, J.; Islam, I.; Mantei, N.; Edwards, Y.;
Biochem. J. 285, 915-923, 1992
A:Title: Sequence of the complete cDNA and the 5' structure of the human sucrase-isomalt
A:Reference number: S24329; MUID:92359963; PMID:1353958
A:Accession: S24329
A:Molecule type: mRNA
A:Residues: 1-661, 'F', 663-931 <CHA>
A:Cross-references: EMBL:X63597
R:Corvel, J.P.; Ferrero, A.; Chambraud, L.; Rigal, A.; Bonicel, J.; Maroux, S.
Gastroenterology 101, 618-625, 1991
A:Title: Expression of sucrase-isomaltase and dipeptidylpeptidase IV in human small inter
A:Reference number: A61136; MUID:91317403; PMID:1677636
A:Accession: A61136
A:Molecule type: protein
A:Residues: 2-14, 'F', 16-20; 1008-1015, 'E', 1017-1021, 'TX', 1024 <GOR>
C:Genetics:
A:Gene: GDB:SI
A:Cross-references: GDB:120377; OMIM:222900
A:Map position: 3q25.2-3q26.2
C:Complex: the two product chains remain associated after cleavage
C:Function: <ISM>


```
QY 1032 LQAGKAEVTGYPPLGTYWYDQVPIEALGSLPPPPAAPPAPREPAIHSGQWVTLPAPLDTIN 1091
Db 697 LKQGA VADAYDPAGNWFDFLNFYSF-AVGG-----DSGKHVLDTPADHVN 741
QY 1092 VHLRAGYIIPLOGPGLTITTESROOPMALAVALTKGGEARGELFWDDGESLVLERCAYTQ 1151
Db 742 VHVREGSIVAMQGEALTTRDARTPYOLLVVASRLNISEGFLDDGELNRMGAGGG--- 798
QY 1152 VIPLARNNTIVNELVAVTSEGAGLQ-----LOKVTVLGATAPQOVLNSG 1196
Db 799 ----NRDNLVPRCVTKGSVLRSEVNVPEVASRKMWSIKGVTFVGFENV-ENVKTYE 853
QY 1197 VPVSNFTYGP-----DTKVLDICV-----SLLMGEQF 1223
Db 854 VRTSERLSPRISLIKTVSDNDPRLFSVEVSKLSLVGKGF 895

RESULT 7
T09143
alpha-glucosidase (EC 3.2.1.20) - spinach
C:Species: Spinacia oleracea (spinach)
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C:Accession: T09143
R:Sugimoto, M.; Furui, S.; Suzuki, Y.
Plant Mol. Biol. 33, 765-768, 1997
A:Title: Molecular cloning and characterization of a cDNA encoding alpha-glucosidase from
A:Reference number: Z16585; MUID:97238484; PMID:9132069
A:Accession: T09143
A>Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-903 <SUG>
A:Cross-references: UNIPROT:O04893; EMBL:D86624; NID:G2081626; PIDN:BAA19924.1; PID:G208
A:Experimental source: strain Dash
C:Superfamily: lysosomal alpha-glucosidase; sucrase/isomaltase homology; trefoil homology
C:Keywords: glycosidase; hydrolase
F:145-799/Domain: sucrase/isomaltase homology <SIM>

Query Match 22.9%; Score 1494.5; DB 2; Length 903;
Best Local Similarity 37.9%; Pred. NO. 5.9e-83;
Matches 340; Conservative 138; Mismatches 296; Indels 123; Gaps 26;

QY 412 YKLENI-----SSSEMGYTA--TLTRTPPTFPFKDILTLRLDVMETENRHLFTKIDPANRR 466
Db 40 YKYSKVNDSGTGRSRTALTALPOLVKNSSVYGP-DIQLLSITASLESNDRLRVITDAKRR 98
QY 467 YEVPLETPRVHRSRAPSPLYSVE-----FSEEPFGV 496
Db 99 WEIPDNILRHQPDPHPHSLSLYLTLSSTPTNRKILLSHPNSDLTFSLINTTTPGF 158
QY 497 IVHRQLDGRVLLNTTVP-----LFFADQFLQLSTSLPS--QYITGLAEHLSPLMLSTS 548
Db 159 TISRKSTHDVLPDPTNPNTFLFIQYHLTLTSSLPGTRAHIYGLGHSKPTFQLAH 218
QY 549 WTRITLWNRDL-APTGALNYGSHPPYLALED---GGSAHGVPFLNSNMDVVQLSPAL 604
Db 219 NOTLTWRAADIPSSNDPNVNLGSHPPYMDVRSSPVAGSTHGVLLNSGMDVEYTG-NRI 277
QY 605 SWRSTGILDVIFLGPESKVVQQLDVVGVYFPMYPYGLFHLRCWGYSTAITRQV 664
Db 278 TYKVIIGDILFFAGSPGVQVEQETRVIGRAPMPYAFGQQCRYGHDVYELQSVV 337
QY 665 ENNTRAHFLPDVQWNDLDYMDSRDFT-----FNKDGPRDFPAMVQELHQQGRRYMYVD 719
Db 338 AGYAKAKIPIEVMMWTDIDYDAYKDFLDPVNPFLDKMKF---VNNLHKGQKVYILD 394
QY 720 PAISSGGPAGSYRYPDEGLRGRVFTNETGQPLIKWFGSTAFDFTNPTALAWEDMV 779
Db 395 PGISTN---KTYETIRGKMDHVL-KRNGKPYLGSVFGVYFDFLKPSSALTFTWDBI 450
QY 780 AEFHDQVDFDGLWIDMNEISNFT---SSPIPGSTLDNPPYKINNNSGVMLPIINKTIPPTA 836
Db 451 KRFLNLLPVDGLWIDMNEISNFT---SSPIPGSTLDNPPYKINNNSGVMLPIINKTIPPTA 508
```

```
QY 837 SHQFLSTYHNHLNVLGLTEATASHALVKARTRPFVISRSTFAGHGRYAGHWTGDVWSS 896
Db 509 MHYGDIPSYNVNHLFGYLEARVTRAAALIKLTKRPFVLRSSTFSGSKYTAHTWGDNAAT 568
QY 897 NEQLASSYPEILQFNLLGVPLVGADVCCPLGNTSEELCVRWOTGLGAFYFPMENHSLLSL 956
Db 569 WNDLYSIPSMDFLGFIPMGADICGFLGNTTEELCRRWJQLGAFYFPSHDHSLGTT 628
QY 957 POEYFSFSEPAQAMRKALTLRYALLPHLYTLFHOAHVAGETVARPLFLFEFFPKDSSTWT 1016
Db 629 YQELRWESVAASA-RKVLGLRYTLFPYILMYEAQLNGPIARPLPFSEFDDDKTYGI 687
QY 1017 DQQLWGEALLITPVLAGKAEVTCYFPLGTWYDL--QTVPIEALGSLPPPPAAPPREPAI 1074
Db 688 SSQFLLGKGMVSPVLKPGVSVTAYFPRGNWDFLFDYTRSVTA----- 731
QY 1075 HSEGOVTLPAPLDTINVHLRAGYIIPLOGFLTTESROOPMALAVALTKGGEARGELF 1134
Db 732 -STGRYVTLSPDPDHINVIQEGNILAMQKAMTTQAAARKTPFFHLVVMNSDCGASFGELF 790
QY 1135 WDDGESLEV-LERGAYTQVIFL---ARNNTIVNELRVVTSEGAGLQ---LOKVTVLGAT 1187
Db 791 LDDGVEVTMGVNRGKWTFKFIAASAKQTCITSDV-VSGEFVAVSKQWIDKVTILG--- 846
QY 1188 APQVLSNGVPVSNFTY-----SPDTK-----VLDIC-VSLLMGSOQF 1223
Db 847 -----LRKGTNGYVTRTGAVTRKDKSKLSTPDRKGEFIVABISGLNLLGREF 898

RESULT 8
JC5463
alpha-glucosidase (EC 3.2.1.20) - sugar beet
N:Alternate names: alpha-D-glucoside glucosylhydrolase
C:Species: Beta vulgaris var. altissima (sugar beet)
C>Date: 17-Jun-1997 #sequence_revision 12-Sep-1997 #text_change 09-Jul-2004
C:Accession: JC5463; PC4330
R:Matsui, H.; Iwanami, S.; Ito, H.; Mori, H.; Honma, M.; Chiba, S.
Biosci. Biotechnol. Biochem. 61, 875-880, 1997
A:Title: Cloning and sequencing of a cDNA encoding alpha-glucosidase from sugar beet.
A:Reference number: JC5463; MUID:97321863; PMID:9178565
A:Accession: JC5463
A:Molecule type: mRNA
A:Residues: 1-913 <MATI>
A:Cross-references: UNIPROT:O04931; DDBJ:D89615; NID:G2217947; PIDN:BAA20343.1; PID:G2191
A:Experimental source: seed; cv. NK-152
A:Accession: PC4330
A:Molecule type: protein
A:Residues: 234-261;310-365;507-541;810-840 <MAT2>
A:Experimental source: seed
C:Comment: This enzyme is an exo-glucosylhydrolase that catalyzes the hydrolysis of alpha-gl
C:Superfamily: lysosomal alpha-glucosidase; sucrase/isomaltase homology; trefoil homology
C:Keywords: glycosidase; hydrolase
F:149-803/Domain: sucrase/isomaltase homology <SIM>

Query Match 22.4%; Score 1458.5; DB 2; Length 913;
Best Local Similarity 37.2%; Pred. NO. 9.4e-81;
Matches 330; Conservative 135; Mismatches 313; Indels 109; Gaps 25;

QY 389 AKQGLQAGMQGPCWCFPPSPYSYKLENLSSEMGYTATLTFTTTPFKOILTLRLDVM 448
Db 32 SKNDQGBAIGYV-----QVKNKVDNSTGKSLTALLQLIRNSPVYGP-DIHFLSFSTAS 85
QY 449 METENRLHFTIKDPANRRYEVPLET-PR-VHSRAPSPLYSVE----- 488
Db 86 FEEDTLIRFTDANNRWEIPNEVLPRPPPPPSLSSLOHLKPKIPQNPQPTTVLSH 145
QY 489 -----FSEEPGVIVHRQLDGRVLLNTTVP-----LFFADQFLQLSTSLPSQ--- 531
Db 146 PHSDLAFTLFTTTPGFTYRKSTHDVLPDPTIPSPNPTTFLIYKDYQLQLSSSLPAQOA 205
QY 532 YITGLAEHLSPLMLSTS WTRITLWNRDLAP--TPGANLVGSHPPYLALEDG---GSAHGVP 587
Db 206 HLUGLGEHTKPTFQLAHNQILTLWNADIASFNRLNLIQSHPPFYMDVRSSPMVGSTHGVF 265
```


C:Accession: T15893
R:Geisel, C.
Submitted to the EMBL Data Library, November 1995
A:Description: The sequence of *C. elegans* cosmid D2096.
A:Reference number: Z18425
A:Accession: T15893
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1743 <GEI>
A:Cross-references: UNIPROT:Q19004; EMBL:U40800; NID:g1065946; PID:g1065949; PIDN:AAA814
C:Genetics:
A:Gene: CESP.D2096.3
A:Introns: 76/1; 100/1; 186/2; 245/3; 384/1; 669/2; 720/2; 737/3; 752/1; 806/2; 862/3; 8

Query Match 21.7%; Score 1416; DB 2; Length 1743;
Best Local Similarity 28.6%; Pred. No. 1e-77;
Matches 385; Conservative 133; Mismatches 433; Indels 334; Gaps 43;

QY 22 SREKQPKSPKRE-----SCGEPRMEKLNQWKAQRLH-----LPPVRL 62
DB 397 SNEHNRGAPTPKRPVFRKERKRTHEEIQ-BELRKIWEIEGIGKHSMSQVSGQFPFKSV 455
QY 63 AELHADLKIQERDELAKWKLDGLDDEGEKARLIRNLNVLAKYGLDGK-----KD 115
DB 456 K-----KDEENE-----DGDQDW-IDHLLARNPRLECRKKTDTNLHRSPDRKS 500
QY 116 ARQVTSNLSGTQEDGLDPRLEKLMHAKRTSGKFSGEELDKLWREFLHHEKVEHYNNVL 175
DB 501 EKVIPAKTRSKILQVQEQPKMKMGICRIDAPSSAEHLNLSLK-----EWELVEDQKLQ 557
QY 176 LETLSRTEIHENIVSPSOLSDIKGVLHSRTELKEKLRN-QGLDLRLRVSHQGYST 234
DB 558 L-SRKAEBELR-----LLH--YTSL--VSSINRKGIOELDGASKKSEIV 596
QY 235 EAEFEPRVIDLWDLAQSANLTDKELEAFR-BELKHFEAKIEKHNHYQKLEIAHEKLRH 293
DB 597 EQEFQK-----RLAEIBORSLEQVRNEKREQWKS-----MDELTEKGEELLAF 640
QY 294 ABSVGDGER-----VSRGREKHA 311
DB 641 FELTSDGGRNDYIWTEDDEDDWYTPYNNHPKSLCSQCAPRQORRIAKREREQAK 700
QY 312 LLEGRTKELGYVKKHQLDLSGRISPARAEFTGAHPRAVP-----TQC----- 358
DB 701 IKVSCLEKEYGIBKRESLRNINCL--KISVVTSPFSLSPKVPVSDQKTLGCPCTFRCT 757
QY 359 ----- 358
DB 758 FSFFQRPVLEFFNFFAENCLHGRKVRKQYPVSHLNIIFLANCKQRLKLMFLKKGSTIV 817
QY 359 -----DVPPNRFDCAPDKAITOQCEARGCCYIPAKQGLQAGQMGQPCFP 405
DB 818 VLLLVFYQNTNSQSDVSKRVDYCPPEGASQACQSRGCIWTEAP---SSSPVGTWCYY 874
QY 406 PSYPSYKLENLSSEMGYATLTRT-----TPFPFKDILTLRLDVMETENLHF 457
DB 875 P-----TESGTVOSTGTNSFVLAAKTKNPGYNNISPLNVKY---SSNGATL 918
QY 458 TIKDAPNRRVYPLETPRVHSRAPSPLYSEF-----SEEPFGVIVHRLDGRVLNNTT 511
DB 919 LUTIGNDDRYVPVNPFPKPSSTE---SLKTSGTIGSDVFSFKVTRASTGVALLWDTIS 975
QY 512 VAPLFFADQFLQLSTSPQYITGLAEHLSPLMLSTWTRITLW---NRDLAPTGA--- 565
DB 976 IGMQFADKFIQIATVLPKSNIVGFDHIK-KIRNLDORYTTPWFARDIGDPSGALS 1034
QY 566 --NLYGSHFPYLALEGGAGHVFLLNSAMDVLOPSPALSWRSTGGILDVIFLGPBP 623
DB 1035 TONLYGVHPFYMCIEADGAGHGVFIILNSNAQEVETGFGPHLLYRTITGGRIDMAFFPGPT 1094
QY 624 KSVVQYLDVGVYPMPPYKGLGFLHCRWYGSSTAITROVVENMTFAHPLDVOQNDLDY 683
DB 1095 EQVNYLOHIGFFLPFAYWALGYQLCRWYGNLDAMKTVISRNQALGIPLDVPYADIDY 1154

QY 684 MDSRDEFTFNKDGFRDPPAMVQELHOGGRRYMMIVDPAISSSGPAGSVRYPYDEGLR---- 739
DB 1155 MNHYEDFT-EGDNWSGFPAYTQOIHAQGLHLIVIFDPAVEV-----DYASFORGINADAS 1208
QY 740 -----RGVFTNETGOLPIKWKWPG-STAFPOPTNP--TALAWMEDWVAEF 782
DB 1209 FIEWARDDQVPHNIQDQYPMKNTKIMLGNVWPDRTAPDFDPRNNTNNAWMEFAQF 1268
QY 783 HDQVPFDGLWIDMNPSPNFIRGSED-----GCP-----NNELENPPYVPVGVV--- 824
DB 1269 HKTLPFDGWDMDMNPSPNFDGTNTVVEEQLASAKLSCFITGSDSLDVPFPYPTQAVYQR 1328
QY 825 -GGTLOAATIC---ASSHQFISTHYNLNLYGLTEAIAHSHALVKARGTRPPIVSRSTFA 880
DB 1329 NGEYLFSTKLCMLGKTAHR-TRDFYDTKNLYGWSERATYQAIPOVTKRSVAVISRSTFP 1387
QY 881 GHGRVAGHWTGDVWSWEQLASVPEILQPNLLGVLCAOVCGFLGNTSEBELCVRWTLQ 940
DB 1388 SSGRYGGHWDGNTARWGDLOTSVIGVMEFNMGVYVVGSDICGPNVGSNEELCLRWHQF 1447
QY 941 GAFYPMFNHNSLLSLPQEPYSFSEPAQQAAMKALTRYALLPHLYTLFHOAHVAGETVA 1000
DB 1448 GAFSFSRDHNSGPNDDQPAWPSVA-NAAKIALTFYFPLFLYSLHYNAARYGHTVI 1506
QY 1001 RPLFLFEPKDSSTVTDHOLLWGEALLITPVLQAGKAEVGYFPPLGTWYDLOTVPIEALG 1060
DB 1507 RPLFPFPKDBETLNISSQFLWGSALMIAPALYQOQTSVHAYFPDSDTWYSLO----- 1558
QY 1061 SLPPPPAAPREPAIHSEGOVW---TLPAPLDTIN-VHLRAGVILIQLOGPLTSTESROOP 1116
DB 1559 -----PETYQKMGFSGFNDVNAPLSLTVPVRGCVLPQSPGTTTTASRLSP 1607
QY 1117 MALAVALTKGGEARGELFWDDGESL 1141
DB 1608 FELLITVKTNAASSGDLYYDGGDDL 1632

RESULT 11
S65057
alpha-glucosidase (EC 3.2.1.20) - barley
C:Species: Hordeum vulgare (barley)
C:Date: 20-Jul-1996 #sequence_revision 27-Feb-1997 #text_change 09-Jul-2004
A:Accession: S65057; S65058
R:Tibbot, B.K.; Skadsen, R.W.
Submitted to the EMBL Data Library, March 1995
A:Description: Molecular cloning and characterization of a gibberellin-inducible alpha-g
A:Reference number: S65057
A:Accession: S65057
A:Molecule type: mRNA
A:Residues: 1-877 <TIB>
A:Cross-references: UNIPROT:Q43763; EMBL:U22450; NID:g944900; PIDN:AA02985.1; PID:g94490
R:Tibbot, B.K.; Skadsen, R.W.
Plant Mol. Biol. 30, 229-241, 1996
A:Title: Molecular cloning and characterization of a gibberellin-inducible, putative alpt
A:Reference number: S65058; MUID:96178863; PMID:8616248
A:Accession: S65058
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 85-115, 'S', 117-121, 123-170, 'D', 172-366, 'G', 368-529, 'H', 531-685, 'G', 687-764 <J
A:Cross-references: EMBL:U22450
C:Superfamily: lysosomal alpha-glucosidase; sucrose/isomaltase homology; trefoil homology
C:Keywords: glycosidase; hydrolase
F:125-769/Domain: sucrose/isomaltase homology <SIM>

Query Match 21.5%; Score 1403.5; DB 2; Length 877;
Best Local Similarity 37.4%; Pred. No. 2e-77;
Matches 331; Conservative 133; Mismatches 327; Indels 95; Gaps 25;

QY 403 CFFPSPSYKLENLSSESMEGYATLT-----RTTTPFPKDIILTLRLDVMETEN 453
DB 14 CLFAPRLCSSKEEGPLAARTVLAIVATMEGALRAEAATGRRSSTGDVQRLAVVASTIDS 73

QY 454 RLHFTIKDPAARRRYEYVPLE-TPR-----VHSRAP-----SPLYSVERS-----EE 492
Db 74 RLVRITDADHPREVPQDIIPRAPGDVLHDAPPASSAPQGRVLSPAGSDLVLTVHAS 133
QY 493 PFGVIVHRLDGRVLLNTTAP--LFFADQFLQLSTLPS--QYITGLAEHLSPMLSTSW 549
Db 134 PFRTVSRRTGTDLT--AGLVFDRKYLEVTSALPAGRASLYGLGHTKTSFRLRH 191
QY 550 TRITLWNRDL--APTPGANLYGSHPPFYALBDGSAHGVLFLNANMDVVLQSPALSWRS 608
Db 192 DSPTLWADIGASVDVNLGSHPPFYMDVRAPGTAHGVLLLSNGMD-VLYGGSYVTKV 250
QY 609 TCGILDVYIFLGPSPKSVQOYLDVGVYPPMPYWGFLHCRWYSSTAITRQVVENMT 668
Db 251 IGVLDYFFYFAGNPLAVALDQYTLTARPAAMPYWSFGFHQCRGYLVNLSDLERVARYA 310
QY 669 RAHFLPDVQNDLDYMSRDTFENKDF--RDPAMVQELHOGGRYMMIVDPAISSSG 726
Db 311 KARIPELVMTDIDYMDGFKDFTLDVRNFTAAELRPFVDRHNAQKYVLLDPGTRVDP 370
QY 727 PAGSYRYPDGLRRGVFITNETQOLIGKVPGSTAPPDFTNPTALAMWDMVAEFHQV 786
Db 371 IDATYGTGVRGMQDIFL-KRNGTNGVGNVPGDVYFPDFMHPAAAEFAWEBSLFRRTI 429
QY 787 PFGLMIDMNEPNSFIRGSDGCPNNELENPY--VPGVVGGLQAATI-CASSHQFLST 843
Db 430 PVDGLMIDMNEISNFNPE----PMNALDDPPYRINNDGTGRPINNKTRVPLAVHYGGVT 485
QY 844 HYNLHNLGYLGEAASHRALVKARGTRPFVSRSTFAGHGRYAGHTWGDVSSWEOLASS 903
Db 486 EYBEHNLFGLEARATGRVLRDTRRRPFVLSRSTFVSGRYTAYTWTGDNAAATGDLRYS 545
QY 904 VPILQFNLGLVPLVGVADVCGFLGNTSEELCVRWTLQGAFFPMRNHNSLLSPQPSYF 963
Db 546 INTMLSFGLGMPMGADICGFGNGTTEELCGRWIQLGAFYPSRDSALFTVRRELYLW 605
QY 964 SEPAQAMRKALTLRVALLPHLYTLFHOAHVAGETVARPLFLFPKDSSTWTDHQLLWG 1023
Db 606 PSVAASG-RKALGRYQLLFYFYLMEAHMTGAPLARPLFFSYPHDVATYGVDRQFLLG 664
QY 1024 EALLITPVLOAGKAEVTGVPPLGTVDLTQVPIEALGSLPPPPAPAPREPAIHSEGQVTL 1083
Db 665 RGVLSFVLEPGPTTVDAYPAGRWRLDYSL-----AVATR-----TGRHVL 709
QY 1084 PAPDLTINVHLRAGYIIPQGLGTTTESRQOPMALAVALKGEARGELFMDGDSLEV 1143
Db 710 PAPADTVNHLTGTLPLQOSALTTSRARTAFHLLVALAEDGTASGYLFLDDGDSPEY 769
QY 1144 LERGATQVIFLARNNTVNE--LVRTSE-----GAGLOLQKVTLG--VATAPOQ 1191
Db 770 GRRSDMSVRF---NYKIPNKGAIKVSEVHNSYAQSRTLVSKVLMGHRSPAAPKK 826
QY 1192 ----VLSNGVPVSNFTYSPDTKVLDI-----CVSLLMGEOF 1223
Db 827 LTVHNSAEVEASS---SAGTRYONAGGLGGVAHIGLSLVGGEF 869

RESULT 12
T38598
Probable family 31 glycosyl hydrolase (alpha glucosidase) precursor - fission yeast (Sch
C:Species: Schizosaccharomyces pombe
C:Date: 20-Oct-2000 #sequence revision 20-Oct-2000 #text_change 09-Jul-2004
C:Accession: T38598; S62559; T38911
R:Pearson, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, November 1995
A:Reference number: Z21801
A:Accession: T38598
A:Molecule type: DNA
A:Residues: 1-385 <PEA>
A:Cross-references: EMBL:Z67961; PIDN:CAA91887.1; GSPDB:GN00066; SPDB:SH
A:Experimental source: strain 972h-; cosmid c30D11
R:Pearson, D.; Churcher, C.M.
submitted to the EMBL Data Library, November 1995

A:Reference number: S62559
A:Accession: S62559
A:Molecule type: DNA
A:Residues: 1-384 <PE2>
A:Cross-references: EMBL:Z67961
R:Pearson, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z21817
A:Accession: T38911
A:Molecule type: DNA
A:Residues: 352-993 <PE3>
A:Cross-references: EMBL:Z67928; PIDN:CAA93572.1; GSPDB:GN00066; SPDB:SPAC56F8.01
A:Experimental source: strain 972h-; cosmid c56F8
C:Genetics:
A:Gene: SPDB:SPAC30D11.01c; SPDB:SPAC56F8.01
A:Map position: 1L
C:Superfamily: Schwanniomycetes glucan 1,4-alpha-glucosidase GAM1; sucrase/isomaltase homo
F:165-384/Domain: sucrase/isomaltase homology (fragment) <SIM>

Query Match 21.2%; Score 1380; DB 2; Length 993;
Best Local Similarity 34.7%; Pred. No. 6.6e-76;
Matches 328; Conservative 142; Mismatches 317; Indels 158; Gaps 28;

QY 408 SYPSYKLENLSSENCYTATLTTRTPFP--KDIILRLDVMMETENRLHFTIKD---- 461
Db 81 SCFGYQARNISYSYGVLAILELAGDACYAYGTDYPLLNVNSYDTEERVHISISDLNOT 140
QY 462 ---PANRR--YEVPLETPRVHGRAPSLYSVSEFPFGVIVHRLDGRVLLNTTVAPLF 516
Db 141 QQLSNRDRVDAPL-FYRSSNFSGNLQNFNTDPPFWITRIADDQVLPTRGNPLI 199
QY 517 FADQFLQLSTLPSQY-ITGLAEHLSPMLSTSWTRITLWNRDLAPTGANLYGSHPPYL 575
Db 200 FEDQVIELTNMVEDYVNYLGSQSQSFLGNLTK-TFWATGYSDSPEANMYGSHPPFY 258
QY 576 A---LEDG-----GSAHGVLNANMDVVLQSPALSWRSTGGTLDVYIFLGP--EPK 624
Db 259 EORYIPIGTNTYTSASHGVLMSSNGMEVLLR-STYIKYRMIGGIIDLFVYSGSTVSPK 317
QY 625 SVVQOYLDVGVYPPMPYMGFLHCRWYSSTAITRQVVENNTRAHFPLDVQNDLDY 684
Db 318 YTIQYVQSIGFTWQPYWSLGFQMSRWGYKTLSDLINRNSYLNASNITEGFMNDIDYM 377
QY 685 DSRRTFTNKDGF-----RDPFAMVOELHOGGRYMMIVDPAISSSGPAGS-----YRPYD 735
Db 378 SEFRFTVNSTAPPENQTLDFRSLDESHQ---HYVPVLDPAIYAANPNKSAADRTYPPY 434
QY 736 EGLRGVFTTNETGQPLICKVWPGSTAFDPFTNPTALAMWDMV-----AEFHQVDP 787
Db 435 SGFEDNIFIKPNPGSAYVGMWPGFVYVPDFTNPAPVLYQWKQIGILNLSTAFGSNYVDLP 494
QY 788 FDGLWIDMNEPNSFIRGSDGCPNNELE---NPPY-VPGVVGGLT----- 828
Db 495 FSGLCIDMNEPISFCIGS---CGSLLKLNLPVHPFSLPGDVNDKNVSYSPEDFNATNTE 551
QY 829 -----QAATICASSHQFLS-----TH----- 844
Db 552 YKSVSRASQSYKATATSEKSHETPSSESLINGKPEFSINYPYALDITDTOTDLAQFGV 611
QY 845 -----YNLHNLGYLGEAASHRALVKAR-GTRPPFVISRSTFAGHGRYAGHTW 891
Db 612 SPNATMHGNTLRNLFNTYGYSEKISFPAALNSIQNIRPPLLSRSTFVSGRYAAHVLG 671
QY 892 DVWSSWEOLASSVPTELQNLGLVPLVGVADVCGFLGNTSEELCVRWTLQGAFFPMRNHN 951
Db 672 DNKSQSDMWSISSILTFENLLGIPMGADVCGYNGTDEELCARMAAGAFLPFRNHN 731
QY 952 SLLSLPQEPYSFSEPAQAMRKALTLRVALLPHLYTLFHOAHVAGETVARPLFLFPKDS 1011
Db 732 SLGSIQEPFPRVASVA-EASRSAIBIRYSLLPYWTLMTHTASVDGTPMVRPLFFEPKQI 790
QY 1012 STWTDHQLLWGEALLITPVLOAGKAEVTGYFPLGT---WYDLQTVPIELAGSLPPPPAA 1068

QY	571	HPFYALEDDGSAHGCVLLNSNAMDVVLQSPALSHRSTGGILDVIFL--GPEPKSVVQ	628
Db	208	HPFYQVRD-GKAGALLMNAHGMVITTEG-RITYKVIGGILDFFAPKSKGNPLSI	265
QY	629	QYLDVVGYFPMPPYWGGLGPHLCRWGYSSTAITRQVVENMTRAHFFPLDVOVNDLNDYMSRR	688
Db	266	NTDILGKPMPSHMLGWHHCYGYPNIDKVTVKYKEANIPLOTVMVDIDYMEETK	325
QY	689	DFTFNKDGFRD--FPAMVQELHOGGRRYMMIVDPAISSGPGAPSYRYPDEGLRGVFTIN	746
Db	326	DFTFDKVNFPQDRMIGLGEQLKDGQNYVMVDPALISAN---TTYBPYVRGTEMDVWIKN	382
QY	747	ETCQPLIGKVPQSTAPPDFTNPTALAWEDMVAEFDQVPPFCGLIMDMNEPSNFRGS-	805
Db	383	ADGSDFIGSVWPGFTTFPDWHPHNPATYWNKEIIDFVDMGLVDGLWDMNEPASFCIGSC	442
QY	806	-----EDGCPNN-----ELB-----NPP-----YVPGVVGTLQ	829
Db	443	GSQKVDAGNQPYRWYTEBEQAANHTWEKELKAMGNPGEERNLLYPKYAINNGAGNLS	502
QY	830	AATICASS-HQFLSTHYNLHNLGLTEAJASHRALVKARG-TRPFVISRSTFAGHGRYAG	887
Db	503	EPTVATTALHYGNIHYDITHNLGHAESHITRQALIKHKINRPPFVLTRSSPPGSKSVG	562
QY	888	HMTGDVSSWEQLASSVPEILQNLGLVPLGVADVCPLGNTSEELCVRTQLGAPYPM	947
Db	563	HMTGDHNSFWPYLKNSIANILNFMFGVSGADVCGFNSDTEELCTRMEIGAFYFPA	622
QY	948	RHNSLSLSPOEPYSFSEPAQAMRKALTLRYALLPHLYTLFHOAVAGETVARPLFLF	1007
Db	623	RHNHNAADQEPYLW-ESTAEASRAINTRYEMLPYFTFLPEENRLGLGVWRPLIFEF	681
QY	1008	PKDSSTWTVDHQLLWGEALLITPVLQAGKAEVGYFPFLGTWYDLQTVPIEALGSLPPPA	1067
Db	682	PAYEELVSDVQTLVGSDDLSPVLDEGTSVKAQFPFGQWYDWTHTV-----	732
QY	1068	APREPAIHSEGQWWTLPAPLDTTNVHLRAGYIIPLOGPGLTTTYESRQOPMALAVALTGG	1127
Db	733	---DNKSNKKVKTVTLDAPLTHIPHIRGGAIIPTKPYTVGTFTATPYNLVIALDKKG	789
QY	1128	EARGELFWDDGESLEVLERGAY	1149
Db	790	QASGRUYIDDGESLEVKSSSGY	811

RESULT 15

JX0281

alpha-2-macroglobulin receptor-associated protein precursor - mouse

N;Alternate names: heparin-binding protein-44

C;Species: Mus musculus (house mouse)

C;Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 09-Jul-2004

C;Accession: JX0281; JX0126

R;Nakamoto, M.; Ozawa, M.; Jacinto, S.D.; Furukawa, T.; Natori, Y.; Shirahama, H.; Yonezawa, J. Biochem. 114, 344-349, 1993

A;Title: Mouse heparin binding protein-44 (HBP-44) associates with brushin, a high-molecular-weight protein kinase. J Biol Chem 268:11025-11031 (1993)

A;Reference number: JX0281; PMID:94110255; PMID:8282724

A;Accession: JX0281

A;Molecule type: mRNA

A;Residues: 1-360 <NAK>

A;Cross-references: UNIPROT:P55302; GB:S67967; NID:G460891; PIDN:AAC60668.1; PID:G460892

R;Furukawa, T.; Ozawa, M.; Huang, R.P.; Muramatsu, T. J. Biochem. 108, 297-302, 1990

A;Title: A heparin binding protein whose expression increases during differentiation of chondrocytes. J Biol Chem 265:11025-11031 (1990)

A;Reference number: JX0126; PMID:91035363; PMID:2229028

A;Accession: JX0126

A;Molecule type: mRNA

A;Residues: 2-360 <FUR>

A;Cross-references: GB:D00622; NID:G220433; PIDN:BAA00500.1; PID:G220434

C;Keywords: chondroitin sulfate proteoglycan; glycoprotein; heparin binding

F;1-37/Domain: signal sequence #status predicted <SIG>

F;38-360/Product: alpha-2-macroglobulin receptor-associated protein #status predicted <M>

F;53,164/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

F;271/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match	20.4%	Score 1327.5;	DB 2;	Length 360;
Best Local Similarity	73.3%	Pred. No. 2.1e-73;		
Matches 253;	Conservative 43;	Mismatches 40;	Indels 9;	Gaps 1;
QY	4	PSGALWLLALLALRTVL-----GSYSREKNQPKSPKRESGEERPMKLNQOLWEKAOR	54	
Db	12	PRQLQLLVLLPLMLVPQPIAGHGKYSREKNPEMAAKRESGEERPMKLNQOLWEKAOR	71	
QY	55	LHLPPVRLAELHADLKIQRDELAWKKLKDGLDEDEGEKEARLIIRNLNVLAKYGLDGKK	114	
Db	72	LHLSPVRLAELHSDLKIQRDELAWKKLKVGLDXDGEKEAKLIHNLNVLARYGLDGRK	131	
QY	115	DARQVTSNLSGTQEDGLDDPRLKLMHKAKTSGKFSGBEELDKLWREFLHHKEKHYENV	174	
Db	132	DAQMVHSNALNEDTQDELGDPRLEKLMHKAKTSGKFSSEELDKLWREFLHHKEKHYENV	191	
QY	175	LLETLSRTEEIHENVISPSDLSIDIKGSVLHSRHTLKEKLRISINQGLDLRLRVSHOGYST	234	
Db	192	LUDTISRABEYENLLSPSDMAHIKSDTLISKHSBELKDRLSINQGLDLRLRVSHOGYST	251	
QY	235	EAEFEEPRVIDLWDLAQSANLTDKLEAFREELKHFEAKIEKHNYQKQLEIAHEKLRHA	294	
Db	252	TTEFEEPRVIDLWDLAQSANLTDKLEAFREELKHFEAKIEKHNYQKQLEISHQKLKHV	311	
QY	295	ESVGDGERVSRREKHALLEGRTKELGYTVKKHLQDLSCRISRAR	339	
Db	312	ESIGDPEHISRNRKKEKYLLEETKELGYTVKKHLQDLSCRISRAR	356	

Search completed: April 5, 2005, 01:07:41
Job time : 58 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2005, 18:57:38 ; Search time 147 Seconds
(without alignments)
4277.777 Million cell updates/sec

Title: US-10-600-862A-7

Perfect score: 6522

Sequence: 1 MRGPGALWLLALRLVLGS.....KVLDTCVSLLMGEQFLVSWC 1228

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4744.5	72.7	952	1 LYAG HUMAN	P10253 homo sapien
2	4730.5	72.5	952	2 Q8IWE7	Q8IWE7 homo sapien
3	4057.5	62.2	937	2 Q9MYM4	Q9MYM4 bos taurus
4	4047	62.1	953	2 Q6P7A9	Q6P7A9 rattus norv
5	4016	61.6	953	1 LYAG_MOUSE	P70699 mus musculu
6	3284	50.4	873	2 Q73632	O73632 coturnix co
7	2573.5	39.5	932	2 Q73626	O73626 coturnix co
8	1904.5	29.2	1856	1 MGA HUMAN	O43451 homo sapien
9	1877.5	28.8	1836	1 SUI5 HUMAN	P14410 homo sapien
10	1875.5	28.8	1826	1 SUI5 RABIT	P07768 oryctolagus
11	1830	28.1	1734	2 Q8TE24	Q8TE24 homo sapien
12	1813.5	27.8	1840	1 SUI5 RAT	P23739 rattus norv
13	1806.5	27.7	1812	1 SUI5 SUNMU	O62653 suncus muri
14	1660.5	25.5	357	1 AMRP_HUMAN	P30533 homo sapien
15	1540.5	23.6	910	2 Q8VW79	Q8VW79 pinus pinas
16	1526.5	23.4	916	2 Q9AVC3	Q9AVC3 physconitre
17	1522.5	23.3	902	2 Q9LTF8	Q9LYF8 arabidopsis
18	1519.5	23.3	902	2 Q22444	O22444 arabidopsis
19	1518	23.3	886	2 Q653V4	Q653V4 oryza sativ
20	1499.5	23.0	885	2 Q653V7	Q653V7 oryza sativ
21	1494.5	22.9	903	1 AGLU SPIOL	O04893 spinacia ol
22	1487	22.8	1053	2 Q876Z7	Q876Z7 mortierella
23	1459.5	22.4	646	2 Q6ZN80	Q6ZN80 homo sapien
24	1459.5	22.4	879	2 Q9LLY2	Q9LLY2 hordeum vul
25	1458.5	22.4	913	1 AGLU BETVU	O04931 beta vulgar
26	1449.5	22.2	907	2 Q9ZP26	Q9ZP26 arabidopsis
27	1449.5	22.2	915	2 Q9S7V7	Q9S7V7 arabidopsis
28	1426	21.9	935	2 Q9ZF04	Q9ZP04 tropaeolum
29	1410.5	21.6	929	2 Q9LGC6	Q9LGC6 oryza sativ
30	1403.5	21.5	877	1 AGLU HORVU	O43763 hordeum vul
31	1402.5	21.5	920	2 Q9NFY8	Q9NFY8 penaeus van

RESULT 1

LYAG_HUMAN
ID LYAG_HUMAN STANDARD; PRT; 952 AA.
AC P10253; Q14351; Q16302;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Lysosomal alpha-glucosidase precursor (EC 3.2.1.20) (Acid maltase).
GN Name=GAA;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE OF 70-89; 123-145; 204-215;
RP 230-249; 332-345; 349-370; 394-409; 480-513; 520-545; 703-719; 726-731
RP AND 795-803.
RC TISSUE=Placenta, Testis, and Urine;
RX MEDLINE=89005058; PubMed=3049072;
RA Hoefslout L.H., Hoogveen-Westerveld M., Kroos M.A., van Beeumen J.,
RA Reuser A.J.J., Oostra B.A.;
RT "Primary structure and processing of lysosomal alpha-glucosidase;
RT homology with the intestinal sucrase-isomaltase complex.";
RT EMBO J. 7:1697-1704(1988).
RL [2]
RN Reuser A.J.J.;
RP REVISIONS.
RL Submitted (JUN-1990) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=90262651; PubMed=2111708;
RA Martiniuk F., Mehler M., Tzall S., Meredith G., Hirschhorn R.;
RT "Sequence of the cDNA and 5'-flanking region for human acid alpha-
RT glucosidase, detection of an intron in the 5' untranslated leader
RT sequence, definition of 18-bp polymorphisms, and differences with
RT previous cDNA and amino acid sequences.";
RL DNA Cell Biol. 9:85-94(1990).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=91097465; PubMed=2268276;
RA Hoefslout L.H., Hoogveen-Westerveld M., Reuser A.J.J., Oostra B.A.;
RT "Characterization of the human lysosomal alpha-glucosidase gene.";
RL Biochem. J. 272:493-497(1990).
RN [5]
RP ACTIVE SITE
RX MEDLINE=91310614; PubMed=1856189;
RA Hermans M.W.P., Kroos M.A., van Beeumen J., Oostra B.A.,
RA Reuser A.J.J.;
RT "Human lysosomal alpha-glucosidase. Characterization of the catalytic
RT site.";
RL J. Biol. Chem. 266:13507-13512(1991).
RN [6]
RP CARBOHYDRATE-LINKAGE SITES.
RX MEDLINE=93168114; PubMed=8435067;

32 1390 21.3 955 2 Q21750
33 1388.5 21.3 969 1 AGLU_SCHPO
34 1385 21.2 1044 2 Q7SHN5
35 1381.5 21.2 936 2 Q19004
36 1380 21.2 993 1 YAJ1_SCHPO
37 1370 21.0 928 2 Q9LEC9
38 1361.5 20.9 995 1 YFZB_SCHPO
39 1352 20.7 992 2 Q9UV08
40 1351 20.7 899 2 Q70I26
41 1341.5 20.6 963 2 Q70I26
42 1330.5 20.4 864 1 AGLU_MUCJA
43 1329.5 20.4 946 1 AMYG_CANAL
44 1327.5 20.4 352 2 Q6PEM5
45 1327.5 20.4 360 1 AMRP_MOUSE

ALIGNMENTS

RA Hermans M.M.P., Wisselaar H.A., Kroos M.A., Oostra B.A.,
RT Reuser A.J.J.;
RA "Human lysosomal alpha-glucosidase: functional characterization of the
RT glycosylation sites";
RL Biochem. J. 289:681-686 (1993).
RN [17]
RN CARBOHYDRATE-LINKAGE SITE ASN-470.
RX MEDLINE=22660472; PubMed=12754519; DOI=10.1038/nbt827;
RA Zhang H., Li X.-J., Martin D.B., Abersold R.;
RA "Identification and quantification of N-linked glycoproteins using
RT hydrazide chemistry, stable isotope labeling and mass spectrometry";
RL Nat. Biotechnol. 21:660-666 (2003).
RN [18]
RN REVIEW ON VARIANTS.
RX MEDLINE=95327152; PubMed=7603530;
RA Reuser A.J.J., Kroos M.A., Hermans M.M.P., Bijvoet A.G.A.;
RA Verbeet M.P., van Diggelen O.P., Kleijer W.J., van der Ploeg A.T.;
RT "Glycogenosis type II (acid maltase deficiency).";
RL Muscle Nerve 3:S61-S69 (1995).
RN [9]
RN VARIANT ASN-91.
RX MEDLINE=90365036; PubMed=2203258;
RA Martinik F., Bodkin M., Tzall S., Hirschhorn R.;
RT "Identification of the base-pair substitution responsible for a human
RT acid alpha glucosidase allele with lower 'affinity' for glycogen (GAA
RT 2) and transient gene expression in deficient cells.";
RL Am. J. Hum. Genet. 47:440-445 (1990).
RN [10]
RN VARIANT GSD-II THR-318.
RX MEDLINE=91353580; PubMed=1652892;
RA Zhong N., Martinik F., Tzall S., Hirschhorn R.;
RT "Identification of a missense mutation in one allele of a patient with
RT Pompe disease, and use of endonuclease digestion of PCR-amplified RNA
RT to demonstrate lack of mRNA expression from the second allele.";
RL Am. J. Hum. Genet. 49:635-645 (1991).
RN [11]
RN VARIANT GSD-II LYS-521.
RX MEDLINE=91379015; PubMed=1898413;
RA Hermans M.M.P., de Graaff E., Kroos M.A., Wisselaar H.A., Oostra B.A.,
RA Reuser A.J.J.;
RT "Identification of a point mutation in the human lysosomal alpha-
RT glucosidase gene causing infantile glycogenosis type II.";
RL Biochem. Biophys. Res. Commun. 179:919-926 (1991).
RN [12]
RN VARIANTS GSD-II ARG-643 AND TRP-725.
RX MEDLINE=94004908; PubMed=8401535;
RA Hermans M.M.P., Kroos M.A., de Graaff E., Oostra B.A., Reuser A.J.J.;
RT "Two mutations affecting the transport and maturation of lysosomal
RT alpha-glucosidase in an adult case of glycogen storage disease type
RT II.";
RL Hum. Mutat. 2:268-273 (1993).
RN [13]
RN VARIANT GSD-II GLU-645, AND VARIANTS ILE-816 AND ILE-927.
RX MEDLINE=93168115; PubMed=8094613;
RA Hermans M.M.P., de Graaff E., Kroos M.A., Wisselaar H.A.,
RA Willemssen R., Oostra B.A., Reuser A.J.J.;
RT "The conservative substitution Asp-645-->Glu in lysosomal alpha-
RT glucosidase affects transport and phosphorylation of the enzyme in an
RT adult patient with glycogen-storage disease type II.";
RL Biochem. J. 289:687-693 (1993).
RN [14]
RN VARIANT GSD-II GLU-645, AND VARIANTS ILE-816 AND ILE-927.
RX MEDLINE=92096118; PubMed=1684505;
RA Martinik F., Mehler M., Bodkin M., Tzall S., Hirschhorn K., Zhong N.,
RA Hirschhorn R.;
RT "Identification of a missense mutation in an adult-onset patient with
RT glycogenosis type II expressing only one allele.";
RL DNA Cell Biol. 10:681-687 (1991).
RN [15]
RN VARIANTS ILE-816 AND ILE-927.
RX MEDLINE=93252406; PubMed=8486380;
RA Hermans M.M.P., Svetkey L.P., Oostra B.A., Chen Y.T., Reuser A.J.J.;
RT "The loss of a polymorphic glycosylation site caused by Thr-927-->Ile
is linked to a second polymorphic Val-816-->Ile substitution in
lysosomal alpha-glucosidase of American blacks.";
RL Genomics 16:300-301 (1993).
RN [16]
RN VARIANT GSD-II VAL-519.
RX MEDLINE=95170739; PubMed=7866409;
RA Huie M.L., Hirschhorn R., Chen A.S., Martinik F., Zhong N.;
RT "Mutation at the catalytic site (M519V) in glycogen storage disease
RT type II (Pompe disease).";
RL Hum. Mutat. 4:291-293 (1994).
RN [17]
RN VARIANT GSD-II TRP-647.
RX MEDLINE=95072571; PubMed=7981676;
RA Huie M.L., Chen A.S., Brooks S.S., Grix A., Hirschhorn R.;
RT "A de novo 13 nt deletion, a newly identified C647W missense mutation
RT and a deletion of exon 18 in infantile onset glycogen storage disease
RT type II (GSDII).";
RL Hum. Mol. Genet. 3:1081-1087 (1994).
RN [18]
RN VARIANT GSD-II LEU-545.
RX MEDLINE=95187163; PubMed=7881422;
RA Hermans M.M.P., de Graaff E., Kroos M.A., Mohkamsing S., Eussen B.J.,
RA Joosse M., Willemssen R., Kleijer W.J., Oostra B.A., Reuser A.J.J.;
RT "The effect of a single base pair deletion (delta 1525) and a C1634T
RT missense mutation (Pro545Leu) on the expression of lysosomal alpha-
RT glucosidase in patients with glycogen storage disease type II.";
RL Hum. Mol. Genet. 3:2213-2218 (1994).
RN [19]
RN VARIANTS GSD-II ARG-299, LYS-903 DEL AND VARIANTS HIS-199; ARG-223 AND
VAL-780.
RX MEDLINE=95233437; PubMed=7717400;
RA Boerkoel C.F., Exelbert R., Nicastri C., Nichols R.C., Miller F.W.,
RA Plotz P.H., Raben N.;
RT "Leaky splicing mutation in the acid maltase gene is associated with
RT delayed onset of glycogenosis type II.";
RL Am. J. Hum. Genet. 56:887-897 (1995).
RN [20]
RN SEQUENCE OF 631-680 FROM N.A., AND VARIANT GSD-II HIS-645.
RX MEDLINE=95209708; PubMed=7695647;
RA Lin C.-Y., Shieh J.-J.;
RT "Identification of a de novo point mutation resulting in infantile
RT form of Pompe's disease.";
RL Biochem. Biophys. Res. Commun. 208:886-893 (1995).
RN [21]
RN VARIANT LYS-689.
RX PubMed=8912788;
RA Huie M.L., Menaker M., McAlpine P.J., Hirschhorn R.;
RT "Identification of an E689K substitution as the molecular basis of the
RT human acid alpha-glucosidase type 4 allozyme (GAA*4).";
RL Ann. Hum. Genet. 60:365-368 (1996).
RN [22]
RN VARIANT GSD-II VAL-529.
RX MEDLINE=96431168; PubMed=8834250;
RA Tsunoda H., Onshima T., Tohyama J., Sasaki M., Sakuragawa N.,
RA Martinik F.;
RT "Acid alpha-glucosidase deficiency: identification and expression of a
RT missense mutation (S529V) in a Japanese adult phenotype.";
RL Hum. Genet. 97:496-499 (1996).
RN [23]
RN VARIANTS GSD-II ASN-645; TRP-647; SER-648; GLN-672 AND TRP-672.
RX MEDLINE=98205825; PubMed=9535769; DOI=10.1006/bbr.1998.8255;
RA Huie M.L., Teujino S., Brooks S.S., Engel A., Elias E., Bonthron D.T.,
RA Bessley C., Shanske S., Dimauro S., Goto Y.I., Hirschhorn R.;
RT "Glycogen storage disease type II: identification of four novel
RT missense mutations (D645N, G648S, R672W, R672Q) and two
RT insertions/deletions in the acid alpha-glucosidase locus of patients
RT of differing phenotype.";
RL Biochem. Biophys. Res. Commun. 244:921-927 (1998).
RN [24]
RN VARIANT GSD-II ARG-309.
RX PubMed=9660056;
RA Kroos M.A., van Leenen D., Verbiest J., Reuser A.J.J., Hermans M.M.P.;
RT "Glycogen storage disease type II: identification of a dinucleotide

RT deletion and a common missense mutation in the lysosomal alpha-glucosidase gene.";

```
Query Match      72.7%; Score 4744.5; DB 1; Length 952;
Best Local Similarity 97.2%; Pred. No. 1.7e-264;
Matches 890; Conservative 3; Mismatches 18; Indels 5; Gaps 1;

QY 313 LGRRTKELGYTVKKHQLQDLSGRISRARABAEAGHPRPRAVPTQCDVPNSRFDCAADK 372
DB 42 LSGSSPVLEETHPAHQGASRPGPR-----DAQAHGPRPRAVPTQCDVPNSRFDCAADK 96

QY 373 AITQEOCEARGCCYIPAKQLOGAQMGPWCFFPPSPYSYKLENLSSEMGTATLTRTT 432
DB 97 AITQEOCEARGCCYIPAKQLOGAQMGPWCFFPPSPYSYKLENLSSEMGTATLTRTT 156

QY 433 PTFFPKDILTLRLDVMMETENRLHFTIKDPANRRYEVPLETPRVHGRAPSPLYSVFSEE 492
DB 157 PTFFPKDILTLRLDVMMETENRLHFTIKDPANRRYEVPLETPRVHGRAPSPLYSVFSEE 216

QY 493 PFGVIVHRLDGRVLLNTTVAFLFFADQFLQSLTSPSOVITGLAEHLSPMLSTWTRI 552
DB 217 PFGVIVHRLDGRVLLNTTVAFLFFADQFLQSLTSPSOVITGLAEHLSPMLSTWTRI 276

QY 553 TLWNRDLATPGANLYGSHPPYLALEDGGSAGHGVFLNNSNAMDVVLQSPALSWRSTGGI 612
DB 277 TLWNRDLATPGANLYGSHPPYLALEDGGSAGHGVFLNNSNAMDVVLQSPALSWRSTGGI 336

QY 613 LDVYIFLGPPEKSVQOYLVDVGPMPYPVWGLGHLCRWGYSTAITQVVENMTTRAHF 672
DB 337 LDVYIFLGPPEKSVQOYLVDVGPMPYPVWGLGHLCRWGYSTAITQVVENMTTRAHF 396

QY 673 PLDVQWNLDDYMSRRDFTFNKGDFRDFPAMVQELHQGGRYVMIVDPAISSSGPAGSYR 732
DB 397 PLDVQWNLDDYMSRRDFTFNKGDFRDFPAMVQELHQGGRYVMIVDPAISSSGPAGSYR 456

QY 733 PYDEGLRGVFTINETGQPLIGKVGFGSTAPPDFTNPTALAWWEDVVAEFDHVPDGLW 792
DB 457 PYDEGLRGVFTINETGQPLIGKVGFGSTAPPDFTNPTALAWWEDVVAEFDHVPDGLW 516

QY 793 IDWNEPSNFRGSEDCGPNNELENPPVPGVVGTTQAATTCASHQFSLTHYNLNLVY 852
DB 517 IDWNEPSNFRGSEDCGPNNELENPPVPGVVGTTQAATTCASHQFSLTHYNLNLVY 576

QY 853 LTAIAASHRALVKARGTRPVIISRSIFAGHGRYAGHWTGDMVSWEQQLASSVPEILQFNL 912
DB 577 LTAIAASHRALVKARGTRPVIISRSIFAGHGRYAGHWTGDMVSWEQQLASSVPEILQFNL 636

QY 913 LGVPLVGADVCGFLGNTSELCVRWTQLGAFYFPMRNHNSLLSLQEPYFSFEPQAQMR 972
DB 637 LGVPLVGADVCGFLGNTSELCVRWTQLGAFYFPMRNHNSLLSLQEPYFSFEPQAQMR 696

QY 973 KALTLYALLPHLYTLFQAHVAGETVARPLFLEFPKDSSTWTDHOLLGREALITPVL 1032
DB 697 KALTLYALLPHLYTLFQAHVAGETVARPLFLEFPKDSSTWTDHOLLGREALITPVL 756

QY 1033 QAGKAEVTVGFFLGTWYDLQTVIEALGSLPPPPAAPREPAIHSEGOVTLPLPDTINV 1092
DB 757 QAGKAEVTVGFFLGTWYDLQTVIEALGSLPPPPAAPREPAIHSEGOVTLPLPDTINV 816

QY 1093 HLRAGYIIPQGPGLTTTSSRQOPMALAVATKXGEARGELFWDGSESLVLRGAYTVQ 1152
DB 817 HLRAGYIIPQGPGLTTTSSRQOPMALAVATKXGEARGELFWDGSESLVLRGAYTVQ 876

QY 1153 IFLARNNTIYNELVRVTSEAGLQKQVTVLGATAPQOVLNGVPSNFTYSPDTKVLD 1212
DB 877 IFLARNNTIYNELVRVTSEAGLQKQVTVLGATAPQOVLNGVPSNFTYSPDTKVLD 936

QY 1213 ICVSLLMGEQFLVSWC 1228
DB 937 ICVSLLMGEQFLVSWC 952
```

RESULT 2

```
Q8IWE7 PRELIMINARY; PRT; 952 AA.
AC Q8IWE7;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE GAA protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Duoenum;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.L., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalek U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
RP SEQUENCE FROM N.A.
RC TISSUE=Duoenum;
RA Strausberg R.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC040431; AAH40431.1; -
DR GO; GO:0004553; F:Hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO; GO:0005975; P:Carbohydrate metabolism; IEA.
DR InterPro; IPR000322; Glyco_hydro_31.
DR InterPro; IPR000519; P_trefoil.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR Pfam; PF00088; Trefoil; 1.
DR SMART; SM00018; PD; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.
DR PROSITE; PS00025; P_TREFOIL; 1.
SQ SEQUENCE 952 AA; 105323 MW; 6E2717BF7201F469 CRC64;
```

Query Match 72.5%; Score 4730.5; DB 2; Length 952;
Best Local Similarity 96.8%; Pred. No. 1.1e-263;
Matches 887; Conservative 4; Mismatches 20; Indels 5; Gaps 1;

```
QY 313 LGRRTKELGYTVKKHQLQDLSGRISRARABAEAGHPRPRAVPTQCDVPNSRFDCAADK 372
DB 42 LSGSSPVLEETHPAHQGASRPGPR-----DAQAHGPRPRAVPTQCDVPNSRFDCAADK 96

QY 373 AITQEOCEARGCCYIPAKQLOGAQMGPWCFFPPSPYSYKLENLSSEMGTATLTRTT 432
DB 97 AITQEOCEARGCCYIPAKQLOGAQMGPWCFFPPSPYSYKLENLSSEMGTATLTRTT 156

QY 433 PTFFPKDILTLRLDVMMETENRLHFTIKDPANRRYEVPLETPRVHGRAPSPLYSVFSEE 492
DB 157 PTFFPKDILTLRLDVMMETENRLHFTIKDPANRRYEVPLETPRVHGRAPSPLYSVFSEE 216

QY 493 PFGVIVHRLDGRVLLNTTVAFLFFADQFLQSLTSPSOVITGLAEHLSPMLSTWTRI 552
DB 217 PFGVIVHRLDGRVLLNTTVAFLFFADQFLQSLTSPSOVITGLAEHLSPMLSTWTRI 276
```

```
QY 553 TLNWRDLAPTPGANLYGSHHPFYALBEGGSAHGVFLNLSNAMDVILQSPSPALSWRSTGGI 612
Db 277 TLNWRDLAPTPGANLYGSHHPFYALBEGGSAHGVFLNLSNAMDVILQSPSPALSWRSTGGI 336
QY 613 LDVYIFLGPEPKSVQOYLDDVCGYDPMPPYVWGLGPHLCRWGYSSTAITRQVVENMTTRAHP 672
Db 337 LDVYIFLGPEPKSVQOYLDDVCGYDPMPPYVWGLGPHLCRWGYSSTAITRQVVENMTTRAHP 396
QY 673 PLDVQWNLDDYMDSRDFTFNKDGFRDFFAMVQELHQGGRYVMYVDPALSSGSPAGSYR 732
Db 397 PLDVQWNLDDYMDSRDFTFNKDGFRDFFAMVQELHQGGRYVMYVDPALSSGSPAGSYR 456
QY 733 PYDEGLRGGVFINETGQPLIGKWPSCSTAFDFTNPTALAWMEDMVAFHDOQVDPGLW 792
Db 457 PYDEGLRGGVFINETGQPLIGKWPSCSTAFDFTNPTALAWMEDMVAFHDOQVDPGLW 516
QY 793 IDWNEPSNFIIRGSEDCPNNELENPPYVGVVGGTLOAATICASHQFSLTHYNLNLHG 852
Db 517 IDWNEPSNFIIRGSEDCPNNELENPPYVGVVGGTLOAATICASHQFSLTHYNLNLHG 576
QY 853 LTEAIAASHRALVKARGTRPFVISRSTFAGHGRVAGHWTGDVWSSWEQLASSVPEILQFNL 912
Db 577 LTEAIAASHRALVKARGTRPFVISRSTFAGHGRVAGHWTGDVWSSWEQLASSVPEILQFNL 636
QY 913 LGVPLVADVCGFLGNTSBEELCVRTQLGAFYFPMRNHNSLSLPQEPYSFSEPAQAMR 972
Db 637 LGVPLVADVCGFLGNTSBEELCVRTQLGAFYFPMRNHNSLSLPQEPYSFSEPAQAMR 696
QY 973 KALTIRYALLPHLYTILFQAHVAGETVARPLFLEFPKDSSTWTDVHQLLWGEALLITPVL 1032
Db 697 KALTIRYALLPHLYTILFQAHVAGETVARPLFLEFPKDSSTWTDVHQLLWGEALLITPVL 756
QY 1033 QAGKAEVTGYFPLGTWYDQVPIEALGSLPPPPAAPREPAHSEGOQWTLPAPLDTINV 1092
Db 757 QAGKAEVTGYFPLGTWYDQVPIEALGSLPPPPAAPREPAHSEGOQWTLPAPLDTINV 816
QY 1093 HLRAGYIIPLOQGLTTTSSRQOPMALAVALTKGGEARGELFWDDGESLEVLERGAYTV 1152
Db 817 HLRAGYIIPLOQGLTTTSSRQOPMALAVALTKGGEARGELFWDDGESLEVLERGAYTV 876
QY 1153 IFLARNNTIVNELRVTSAGLQLOKVTVLGATAPQOVLNGVPSNFTYSPDTKVLD 1212
Db 877 IFLARNNTIVNELRVTSAGLQLOKVTVLGATAPQOVLNGVPSNFTYSPDTKVLD 936
QY 1213 ICVSLLMGEQFLVSWC 1228
Db 937 ICVSLLMGEQFLVSWC 952

RESULT 3
Q9MTM4
ID Q9MYM4 PRELIMINARY; PRT; 937 AA.
AC Q9MYM4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Acidic alpha-glucosidase.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20188772; PubMed=10723725;
RA Dennis J.A., Moran C., Healy P.J.;
RT "The bovine alpha-glucosidase gene: coding region, genomic structure,
RL and mutations that cause bovine generalized glycosenosis.",
RL Mamm. Genome 11:206-212(2000).
DR EMBL; AF171666; AAF81637.1; -.
DR EMBL; AF171665; AAF81636.1; -.
DR HSSP; P01359; 2SP.
DR Go; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
```

```
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR00322; Glyco_hydro_31.
DR InterPro; IPR00519; P_trefoil_1.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR Pfam; PF00088; Trefoil_1.
DR SMART; SM00016; PD; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.
DR PROSITE; PS00025; P_TREFOIL; UNKNOWN 1.
SQ SEQUENCE 937 AA; 104756 MW; B7E9A2D66BAF5026 CRC64;

Query Match 62.2%; Score 4057.5; DB 2; Length 937;
Best Local Similarity 85.0%; Pred. No. 6.2e-225;
Matches 747; Conservative 60; Mismatches 69; Indels 3; Gaps 2;

QY 349 GRPRAVPTQCDVPPNSRFDCAKDAITOECEARGCCYIPAKOGLOGAQMGQPMCFPPPS 408
Db 61 GSFRAAPTCQCDLPPNSRFDCAKDKGITQOCEARGCCYMPA-EWPPDAQMGQPMCFPPPS 119
QY 409 YPSYKLENLSSEMGYTATLTRTPTTFFPKOITLRLLDMMETENRLHFTIKDPANRRYE 468
Db 120 YPSYLENLTITETGCTATLTRAVTFFPKOITLRLLDMMETESRLHFTIKDPANRRYE 179
QY 469 VPLETPRVHSRAPSPLYSVFSEEPFGVIVHQLDGRVLLNTTVAPLFPADQFLQLSTSL 528
Db 180 VPLETPRVYSQAPFTLYSVFSEEPFGVIVRKLDRVLLNTTVAPLFPADQFLQLSTSL 239
QY 529 PSQYITGLAEHLSPMLSTSWTRITLWNRDLAPTGANLYGSHHPFYALBEGGSAHGVFL 588
Db 240 PSQHITGLAEHLGSLMLSTNWTKITLWNRDLAPENVNLYGSHHPFYALBEGGSAHGVFL 299
QY 589 LNSNAMDVVLQSPSPALSWRSTGGILDVYIFLGPEPKSVVQOYLDDVGVVPMFPYMGGLGFH 648
Db 300 LNSNAMDVVLQSPSPALSWRSTGGILDVYIFLGPEPKSVVQOYLDDVGVVPMFPYMGGLGFH 359
QY 649 LCRWGYSSATITRQVVENMTTRAHPFLDVQWNLDDYMDSRDFTFNKDGFRDPPAMVQELH 708
Db 360 LCRWGYSTSAITRQVVENMTTRAHPFLDVQWNLDDYMDARRDFTFNKDHFGDPPAMVQELH 419
QY 709 QGGRYVMYVDPALSSGSPAGSYRYPVDEGLRGVFTITNETGQPLIGKWPSCSTAFDFTN 768
Db 420 QGGRYVMYVDPALSSGSPAGSYRYPVDEGLRGVFTITNETGQPLIGKWPSCSTAFDFTN 479
QY 769 PTALAWMEDVAFEDHQPFDGLWIDMNEPSNFIIRGSEDCPNNELENPPYVGVVGGTL 828
Db 480 PETLDWQDMVTEPHAQVFPDGMWIDMNEPSNFIIRGSEDCPNNELENPPYVGVVGGTL 539
QY 829 QAATICASHQPLSTHYNLNLHGTLTEALASHRALVKARGMRPFVISRSTFAGHGRYAGH 888
Db 540 RAAATICASHQPLSTHYNLNLHGTLTEALASHRALVKARGMRPFVISRSTFAGHGRYAGH 599
QY 889 WTGDVWSSWEQLASSVPEILQFNLGVLGADVCGFLGNTSEELCVRTWQLGAFVPPMR 948
Db 600 WTGDVWSSWEQLASSVPEILQFNLGVLGADVCGFLGNTSEELCVRTWQLGAFVPPMR 659
QY 949 NNSLSLSPQEPYSFSEPAQAMRKALTRVALLPHLYTLFQAHVAGETVARPLFLEPP 1008
Db 660 NENALNSQEPYRSETAQAMRKAFTRVLLPYLYTLFRAHVRGETVARPLFLEPP 719
QY 1009 KDSSTWTDVHQLLWGEALLITPVLQAGKAEVTGYFPLGTWYDQVPIEALGSLPPPPAA 1068
Db 720 EDPSTWTDVDRQLLWGEALLITPVLAEKVEVTGYFPGTQWYDQVPIEALGSLPPPPAA 777
QY 1069 PREPAIHSEGOQWTLPAPLDTINVHLRAGYIIPLOQGLTTTSSRQOPMALAVALTKGGE 1128
Db 778 PUTSVHSGQWTLPAPLDTINVHLRAGYIIPLOQGLTTTSSRQOPMALAVALTKGGE 837
QY 1129 ARGELFWDDGESLEVLERGAYTVQIFLARNNTIVNELRVTSAGLQLOKVTVLGATAP 1188
Db 838 AQGELFWDDGESLEVLERGAYTVQIFLARNNTIVNELRVTSAGLQLOKVTVLGATAP 897
QY 1189 PQOVLNGVPSNFTYSPDTKVLDICVSLLMGEQFLVSW 1227
```



```

RT lysosomal alpha-glucosidase.";
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J, and NOD; TISSUE=Heart, and Thymus;
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nakaide I., Osato N., Saito K., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Balderelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Reisel K.W.,
RA Blake J.A., Bradt D., Brusci V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmer S., Gustinchik S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasava Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Malcais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wegner L., Wohlschlag C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Inotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shingawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Breast tumor;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Henken C.M., Schuler G.D.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Schetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinska M.I., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: Essential for the degradation of glycogen to glucose in
CC lysosomes.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-
CC linked D-glucose residues with release of D-glucose.
CC -!- SUBCELLULAR LOCATION: Lysosomal.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 31 family.
CC -!- SIMILARITY: Contains 1 P-type (trefoil) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/

```

```

or send an email to license@isb-sib.ch).
-----
CC EMBL; U49351; AAB06943.1; -.
CC DR EMBL; AK052211; BAC34888.1; -.
CC DR EMBL; AK088481; BAC40382.1; -.
CC DR EMBL; BC010210; AAH10210.1; -.
CC DR HSP; F01359; 2PSP.
CC DR MGD; MGI:95609; Gaa.
CC DR GO; GO:0004558; F:alpha-glucosidase activity; IMP.
CC DR GO; GO:0005980; P:glycogen catabolism; IMP.
CC DR InterPro; IPR000322; Glyco_hydro_31.
CC DR InterPro; IPR000519; P_trefoil.
CC Pfam; PF01055; Glyco_hydro_31; 1.
CC Pfam; PF00088; Trefol; 1.
CC SMART; SM00018; PD; 1.
CC PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
CC PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.
CC PROSITE; PS00025; P_TREFOIL; 1.
CC Glycoprotein; Glycosidase; Hydrolase; Lysosome; Signal.
CC SIGNAL 1 27
CC PROPEP 28 69 By similarity.
CC CHAIN 70 953 Lysosomal alpha-glucosidase.
CC DOMAIN 81 130 P-type.
CC ACT_SITE 518 518 Nucleophile (By similarity).
CC ACT_SITE 521 521 By similarity.
CC DISULFID 82 109 By similarity.
CC DISULFID 92 108 By similarity.
CC DISULFID 103 127 By similarity.
CC CARBOHYD 140 140 N-linked (GlcNAc...) (Potential).
CC CARBOHYD 233 233 N-linked (GlcNAc...) (Potential).
CC CARBOHYD 390 390 N-linked (GlcNAc...) (Potential).
CC CARBOHYD 470 470 N-linked (GlcNAc...) (Potential).
CC CARBOHYD 883 883 N-linked (GlcNAc...) (Potential).
CC CARBOHYD 926 926 N-linked (GlcNAc...) (Potential).
CC CARBOHYD 933 933 N-linked (GlcNAc...) (Potential).
CC CONFLICT 62 62 K -> E (in Ref. 1).
CC CONFLICT 254 254 S -> A (in Ref. 1).
CC CONFLICT 430 431 EL -> DV (in Ref. 1).
CC CONFLICT 434 434 D -> G (in Ref. 1 and 3).
CC CONFLICT 481 481 W -> C (in Ref. 1).
CC CONFLICT 615 615 G -> E (in Ref. 1).
CC CONFLICT 619 619 S -> T (in Ref. 1).
CC CONFLICT 732 732 P -> R (in Ref. 1).
CC CONFLICT 777 777 M -> V (in Ref. 1 and 3).
CC CONFLICT 871 871 R -> H (in Ref. 1).
CC CONFLICT 903 903 R -> K (in Ref. 1).
CC SEQUENCE 953 AA; 106247 MW; 956B89685FB5FF81 CRC64;
Query Match 61.6%; Score 4016; DB 1; Length 953;
Best Local Similarity 80.7%; Pred. No. 1.6e-222;
Matches 747; Conservative 69; Mismatches 18; Gaps 4;
QY 320 LGYTKKHL-----QDL-----SGRISRAAEATGAHPG-----RPRVPTQCDVPP 362
DB 27 LGHLMRLMLLPQDLHSSGLWKYTPHQGKYPGLHIOEQTEQKEAPTQCDVPP 86
QY 363 NSRFDCAPDKAITQRCQCEARGCCYIPAKQGLQGAQMGQPCFPFSPYSYKLENSSSM 422
DB 87 NSRFDCAPDKGISQCEARGCCYIPAGQVLKEPQIQGQPCFPFSPYSYKLENSSTES 146
QY 423 GYATLTTRTTTFFPKDILTLDVMMETENLHTIKDPANRRYEVLETRVHSRAPS 482
DB 147 GYATLTTRTTTFFPKDILTLDVMMETENLHTIKDPANRRYEVLETRVHSRAPS 206
QY 483 PLYSVFSEEPFGVIVRHQDGRVLLNTTVPFLPADQFLQSLTSLPSQYITGLAEHLSP 542
DB 207 PLYSVFSEEPFGVIVRHQDGRVLLNTTVPFLPADQFLQSLTSLPSQYITGLAEHLSP 266
QY 543 LMLSTSWTRITLWNRDLAPTGANLYGSHPPFYLALEDGGAHGVFLLSNAMDVVLQSP 602
DB 267 LMLSTDMARITLWNRDTPPSQGTNLYGSHPPFYLALEDGGAHGVFLLSNAMDVVLQSP 326
QY 603 ALSWRSTGILDVILFLFPGEPKSVVQVQVLDVVGVYFPMPPYWGFLGHLCRWGSSTAITRQ 662

```



```
Db 327 ALTWRSTGILDDVYVFLGPEKPSWQOYLDVGVGPPMPYWGFLPFLCRWGSSTAIVRQ 386
Qy 663 VVENMTRAFPLDQVNDLDYMSRDDFTFNKDGFRDFFAMVQELHQGGRYMWIVDPAI 722
Db 387 VVENMTRTFPLDQVNDLDYMDARDFTFNQDSFADFFDMVRELHQDGRYMWIVDPAI 446
Qy 723 SSGPAGSYRYPDEGLRRGVFINETGQPLIGKWPFGSTAFDFTNPTALAMWEDVAF 782
Db 447 SSGPAGSYRYPDEGLRRGVFINETGQPLIGKWPFGSTAFDFTNPTALAMWEDVAF 506
Qy 783 HQVPPDGLWIDMNEPSNIRGSEDCPNNELENPPYVGVGGTLOAATICASSHQFLS 842
Db 507 HQVPPDGLWIDMNEPSNIRGSEDCPNNELENPPYVGVGGTLOAATICASSHQFLS 566
Qy 843 THYNLNLGLTEAIAASHALVKARTRPFVSRSTFAGHGRVAGHWTGDMVSSWQOLAS 902
Db 567 THYNLNLGLTEAIAASHALVKARTRPFVSRSTFAGHGRVAGHWTGDMVSSWQOLAS 526
Qy 903 SYPEILOFNLGLVPLGVADVCGFLGNTSEELCVRWMTOLGAFYPMFNNHNSLLSLPQEPS 962
Db 627 SYPEILOFNLGLVPLGVADVCGFLGNTSEELCVRWMTOLGAFYPMFNNHNSLLSLPQEPS 686
Qy 963 FSEPAQOAMRKALTYRALLPHLYTLFHOAHVAGETVARPLFLEPKDSSTWVDHQLLW 1022
Db 687 FSEPAQOAMRKALTYRALLPHLYTLFHOAHVAGETVARPLFLEPKDSSTWVDHQLLW 746
Qy 1023 GRALLTTPVLOAGKAEVTVGPELGTWYDLOVPIEALGSLPPEPPAAR-EPATHSGQWV 1081
Db 747 GRALLTTPVLOAGKAEVTVGPELGTWYDLOVPIEALGSLPPEPPAAR-EPATHSGQWV 806
Qy 1082 TLPAPLDTINVHLRAGYIIPLOQPGITTTTESRQOPMALAVALTKGGEARGELFWDGDES 1141
Db 807 TLPAPLDTINVHLRAGYIIPLOQPGITTTTESRQOPMALAVALTKGGEARGELFWDGDES 866
Qy 1142 EVLERGAYTQVPLARNTIIVNELVTVTSEGAGLQLOKVTLVGVATAPQOVLNGVPSN 1201
Db 867 AVLERGAYTQVPLARNTIIVNELVTVTSEGAGLQLOKVTLVGVATAPQOVLNGVPSN 926
Qy 1202 FTVSPDTKVLIDICVSLMGEQPLVSW 1227
Db 927 FTVSPDTKVLIDICVSLMGEQPLVSW 952

RESULT 6
073632 PRELIMINARY; PRT; 873 AA.
AC 073632;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Acid alpha glucosidase.
GN Name=gAaii; Synonyms=gaa2;
OS Coturnix coturnix japonica (Japanese quail).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Coturnix.
OX NCBI_TaxID=93934;
RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=98201735; PubMed=9540858; DOI=10.1016/S0925-4439(97)00092-6;
RA Kunita R., Nakabayashi O., Wu J.Y., Hagiwara Y., Mizutani M.,
RA Pennybacker M., Chen Y.T., Kikuchi T.;
RT "Molecular cloning of acid alpha-glucosidase cDNA of Japanese quail
RT (Coturnix coturnix japonica) and the lack of its mRNA in acid maltase
RT deficient quails";
RL Biochim. Biophys. Acta 1362:269-278(1997).
RN [2]
RN SEQUENCE FROM N.A.
RP Nakabayashi O.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB006754; BAA25890.2; -
DR EMBL; AB081290; BAC15596.1; -
```

```
DR HSP; Q07654; IE9T.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR000322; Glyco_hydro_31.
DR InterPro; IPR000519; P_trefoil.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR SMART; SM00088; Trefoil; 1.
DR SMART; SM00018; PD; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
SQ SEQUENCE 873 AA; 96696 MW; CEADAFE3ACFCBD CRC64;

Query Match 50.4%; Score 3284; DB 2; Length 873;
Best Local Similarity 66.7%; Pred. No. 2e-180;
Matches 587; Conservative 117; Mismatches 156; Indels 20; Gaps 2;

Qy 348 PCRPRAPVTCQCDVPPNRFDCAPDKRAITQECCEARCCYIPAKQGIQGAQMGPWCFFPP 407
Db 13 PSASSASDACSLLPDERDFDCGERLLARADCEARCCYAPSGSGSG--GPPWCFFPL 69

Qy 408 SYPSYKLENLSSEMGYTATLTRTTPFPKDLILRLDMMETENRLHFTIKDPANRRY 467
Db 70 GYRSYRADNVATAGYSARLRVVPFLPADVGTILRLDVMETESRLRFTPRDPARQY 129

Qy 468 EYPLTTPRVHSPAPSLYSVESEBPPGVIHROLDGRVILNTTVAIPFADQFLQIS 527
Db 130 EYPMATPRVSTRAADTLXGVQLQDPFGIVVFRQDGVLLNTSVAPLFPADQFLQIS 189

Qy 528 LPSOYITGLAEHLSPILMLSTSWTRITLWNRDLAPTPGANLYGSHPPYLALEDGSGAHGV 587
Db 190 LPSRISFISGLERLAPLILDTAWTKLWNRDNAPAPQVNLGSHPPYLVLEDGSGAHGV 249

Qy 588 LLNSNAMDVLQSPALSNRSTGGILDVYIFLGPEPKSVVQOYLDVGVGPPMPYWG 647
Db 250 LLNSNAMDVLQSPALSNRSTGGILDVYIFLGPEPKSVVQOYLDVGVGPPMPYWG 309

Qy 648 HLCRGYSSTAITRQVVENNTAHPFLDVQWNDLDYMSRRDFTFNKDFRDPAMVQEL 707
Db 310 HLCRGYSSTAITRQVVENNTAHPFLDVQWNDLDYMSRRDFTFNKDFRDPAMVQEL 369

Qy 708 HCGRRYMWIVDPAISSGPAGSYRYPDEGLRRGVFINETGQPLIGKWPFGSTAFDFT 767
Db 370 HCGRRYMWIVDPAISSGPAGSYRYPDEGLRRGVFINETGQPLIGKWPFGSTAFDFT 429

Qy 768 NPTALAMWEDVAFHDDQVFDGLWIDMNEPSNIRGSEDCPNNELENPPYVGVGGT 827
Db 430 NPTALAMWEDVAFHDDQVFDGLWIDMNEPSNIRGSEDCPNNELENPPYVGVGGT 489

Qy 828 LOAATICASSHQFLSTHYNLNLGLTEAIAASHALVKARTRPFVSRSTFAGHGRYAG 887
Db 490 LOAATICASSHQFLSTHYNLNLGLTEAIAASHALVKARTRPFVSRSTFAGHGRYAG 549

Qy 888 HWTGDMVSSWEOLASSVPEILOFNLLGVPLGVADVCGFLGNTSEELCVRWMTOLGAFY 947
Db 550 HWTGDMVSSWEOLASSVPEILOFNLLGVPLGVADVCGFLGNTSEELCVRWMTOLGAFY 609

Qy 948 RHNHNSLLSQPSPYSPFSPQAQAMRKALTYRALLPHLYTLFHOAHVAGETVARPLF 1007
Db 610 RHNHNSLLSQPSPYSPFSPQAQAMRKALTYRALLPHLYTLFHOAHVAGETVARPLF 669

Qy 1008 PKDSSTWVDHQLLWGEALLITPVIQAGKAEVTVGPELGTWYDLOVPIEALGSLP 1067
Db 670 PKDSSTWVDHQLLWGEALLITPVIQAGKAEVTVGPELGTWYDLOVPIEALGSLP 713

Qy 1068 APREPAIHSEGOVTLPAPLDTINVHLRAGYIIPLOQPGITTTTESRQOPMALAVALTK 1127
Db 714 -TGDSTIHSKGMWILLAAPLDTINVHLRAGYIIPLOQPGITTTTESRQOPMALAVALTK 772

Qy 1128 EARGELFWDDGESLVEVGAYTVIFLARNTIIVNELVTVTSEGAGLQLOKVTLVGVAT 1187
Db 773 EARGELFWDDGESLVEVGAYTVIFLARNTIIVNELVTVTSEGAGLQLOKVTLVGVAT 832

Qy 1188 APQOVLNGVPSNFTYSPDTKVLIDICVSLMGEQFLVSW 1227
Db 1188 APQOVLNGVPSNFTYSPDTKVLIDICVSLMGEQFLVSW 1227
```

```

Db      833 APQRLVANGIPVEDFSYRSDTQVLRVSVSLPMWEQFVPEW 872
RESULT 7
O73626 PRELIMINARY; PRT; 932 AA.
AC O73626;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE DE alpha glucosidase.
OS Name=GAAL; Synonyms=Gaal;
GN Coturnix coturnix japonica (Japanese quail).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Coturnix.
OX NCBI_TaxID=93934;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=98201735; PubMed=9540858; DOI=10.1016/S0925-4439(97)00092-6;
RA Kunita R., Nakabayashi O., Wu J.Y., Hagiwara Y., Mizutani M.,
RA Pennybacker M., Chen Y.T., Kikuchi T.;
RT "Molecular cloning of acid alpha-glucosidase cDNA of Japanese quail
RT (Coturnix coturnix japonica) and the lack of its mRNA in acid maltase
RT deficient quails.";
RL Biochim. Biophys. Acta 1362:269-278(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Nakabayashi O.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB000967; BAA25884.1; -
DR EMBL; AB081289; BAC15595.1; -
DR HSSP; Q07654; 1E9T.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR00322; Glyco_hydro_31.
DR InterPro; IPR000519; P_trefoil.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR Pfam; PF00088; Trefoil; 1.
DR SMART; SM00018; PD; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.
SQ SEQUENCE 932 AA; 104689 MW; B62E182F03DE3F61 CRC64;
Query Match 39.5%; Score 2573.5; DB 2; Length 932;
Best Local Similarity 53.0%; Pred.No.1.8e-139;
Matches 472; Conservative 144; Mismatches 248; Indels 27; Gaps 9;
QY 341 EAETGAHGRPRVPTQCD--VPNSRFDCAPOK--AITQCEARGCCYIPAKOGLQGAQ 397
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
64 QVSRGWHGPAP---PAQCHLVESHRYDCYPERNVVVTOELCESRGCCFIQTLPVAV-GGK 119
QY 398 MGQPCWFFPPSPYSKLENSSEMCGYATLRTTPTFPFKDILTURLDMMETENRLHF 457
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
120 RGVPCFYPDPFPSPYVQSLNQTVLGMLVRRERAKYVPKDIQMLRMDVETQNTRLHI 179
QY 458 TIKDPAARREYVPLETPRVHSRAPSLYSVESEEPGVIHRLDGRVLLNTTAPLPF 517
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
180 KITDAANRYEVPLEVPVRVTKAENPTYSLSEISQDPFGVLLRRRGQTGVLLNTTAPLPF 239
QY 518 ADQFLQSLTSLPSQYITGLAEHLSPMLSTSTRTITLWNRDLAPTGPANLYGSHPPFYAL 577
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
240 ADQFLQISITLPSRFYGLGEHRSTLLHSLDNTLTLMWARDVAPTESFNLYGAHPFYLLM 299
QY 578 EGGSGHGVFLNSNAMDVVLSPSPALSWRSTGGILDVIFIGPEPKSVVQQYLDVVGYP 637
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
300 BEGGDAHGVLNLSNAMEVALOPAPGLTWRITGGVLDYIFILGPDPMVVIQQYQVEIGFP 359
QY 638 FPPPYWGLGFLHLCRWGYSSTAIQVVENNTRAHPPLDQWMDLSDYMDSRDFTFNKQGF 697
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
360 AMPPLWALGFLHLCRWGYSNETWQTAKARNFQIPQDAQWMDIDYMDGRDFTFPQPF 419

```

```

QY 698 RDPFAMVQELHQGRYRMYMIVDPAISSSGPAGSYRYPDEGLRGRVFTINTEQPLIGKW 757
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
420 ASLPSLVEDLHKHGQHYIILDPGISSTSPRGSYWPFDGLRGLFLNTTQGTLLGQVW 479
QY 758 PGSTAPDPDFTPTALAWEDMVAEPHDQVPPDGLMIDMNEPSNPIRSGSDGCPNHELENP 817
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
480 PGYTAYPDPSNTDTHQWLENLQRFHTHPVFDGLMIDMNEPSNFMDSGSEGGPCGELDSP 539
QY 818 PVPGVGVGTLLQAATTCASSHQFLSTHYNLNHLNLYGLTEAIAASHRALVARGTRPPFVISR 877
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
540 PYPYAVLGNLSUTAKTVCASAEQNASVYHNHLNLYGLKEAEATASALIRIRGRPPFVISR 599
QY 878 TFAGHGRYAGHWITGDVWSSWEQASVPEILQFNLLGVPLVGVADYCGFLGNTSEELCVRW 937
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
600 TFPQSGRYSGHWLGNRSQWDMYYSIFGMLSFSFGIPLVGVADICGFSGSTSEELCTRW 659
QY 938 TOLGAFYPPMHNHLSLSPQEPYSPFPAQAQMRKALTLYRALLPHLYTLFPHQAHVAGE 997
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
660 MOLGAFYPPSRNHNQNEKAQDPTAFSPSARTAMKDALLTRYSLLPFLYTLFPHRAHQE 719
QY 998 TVARPLFLFEFPKDSSTWTVHQLLWGEALLTPVLQAGKAEVTPFPPLGTWYDLQVPIE 1057
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
720 TVARPLFEFFPDWVATYGLDRQFLMGQSLLTVPVLEPGADSVLGYFPQGVWYDFYT---- 775
QY 1058 ALGSLPPPPAAPREPAIHSEGOVTLPAPLDTINVHLRAGYIIPLOGPLTTTESRQPM 1117
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
776 --GS-----SVNSSGEMLKSLAPLDHLNHLREGSILPTQKPGITSKATRGNDL 822
QY 1118 ALAVALTGGEGARGELFWDDGESLEVLERGATQVIFLARNTIYNELVRVTSEGAGLOL 1177
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
823 HLIVALTSTRATAWGDLFWDDGESLTFFEGNSYLVFNATENIFTSNVLHASTEATDVTI 882
QY 1178 QKVTVLGVATAPQOVLVSNQVPSVNTSPDTKVLDIC--VSLLMGBOFLVSW 1227
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
883 DAVSYFGVQEPSPKVLDDG--QEKPSYL-DNQVLTVSLGLVLSQGSFLQW 931
RESULT 8
MGA HUMAN
ID MGA_HUMAN STANDARD; PRT; 1856 AA.
AC O4351;
DT 30-MAY-2000 (Rel. 39, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Maltase-glucosylase, intestinal (Includes: Maltase (EC 3.2.1.20)
DE (Alpha-glucosidase); Glucoamylase (EC 3.2.1.3) (Glucan 1,4-alpha-
DE glucosidase)].
GN Name=MGA; Synonyms=MGA, MGAML;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Small intestine;
RX MEDLINE=98112863; PubMed=9446624; DOI=10.1074/jbc.273.5.3076;
RA Nichols B.L., Eldering J.A., Avery S.E., Hahn D., Quaroni A.,
RA Sterchi E.E.;
RT "Human small intestinal maltase-glucoamylase cDNA cloning. Homology to
RT sucrase-isomaltase.";
RL J. Biol. Chem. 273:3076-3081(1998).
RN [2]
RP REVISIONS TO 776; 1049; 1100; 1541; 1612 AND 1811.
RA Nichols B.L., Eldering J.A., Avery S.E., Hahn D., Quaroni A.,
RA Sterchi E.E.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP CHARACTERIZATION.
RC TISSUE=Small intestine mucosa;
RX MEDLINE=89066802; PubMed=3143729;
RA Naim H.Y., Sterchi E.E., Lentze M.J.;
RT "Structure, biosynthesis, and glycosylation of human small intestinal
RT maltase-glucoamylase.";

```

J. Biol. Chem. 263:19709-19717(1988).

[4] SULFATION.

RL MEDLINE=88082658; PubMed=3121301;

RX Danielsen E.M.;

RA "Tyrosine sulfation, a post-translational modification of microvillar

RT enzymes in the small intestinal enterocyte.";

RL EMBO J. 6:2891-2896(1987).

CC -!- FUNCTION: May serve as an alternate pathway for starch digestion

CC when luminal alpha-amylase activity is reduced because of

CC immaturity or malnutrition. May play a unique role in the

CC digestion of malted dietary oligosaccharides used in food

CC manufacturing.

CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-

CC linked D-glucose residues with release of D-glucose.

CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-

CC glucose residues successively from non-reducing ends of the chains

CC with release of beta-D-glucose.

CC -!- SUBUNIT: Monomer.

CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Brush border.

CC -!- TISSUE SPECIFICITY: Expressed in small intestine, granulocyte, and

CC kidney but not in salivary gland or pancreas.

CC -!- PTM: N- and O-glycosylated.

CC -!- PTM: Does not undergo intracellular or extracellular proteolytic

CC cleavage.

CC -!- PTM: Sulfated (By similarity).

CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 31 family.

CC -!- SIMILARITY: Contains 2 P-type (trefoil) domains.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AF016833; AAC39568.2; -.

DR Genbank; HGNC:7043; MGAM.

DR MIM; 154360; -.

DR GO; GO:0003824; F:catalytic activity; TAS.

DR GO; GO:0003983; P:starch catabolism; TAS.

DR InterPro; IPR000322; Glyco_hydro_31.

DR InterPro; IPR000519; P_trefoil.

DR Pfam; PF01055; Glyco_hydro_31; 2.

DR Pfam; PF00088; Trefoil; 2.

DR SMART; SM00018; PD; 2.

DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 2.

DR PROSITE; PS00025; P_TREFOIL; 1.

KW Direct protein sequencing; Glycoprotein; Glycosidase; Hydrolase;

KW Multifunctional enzyme; Repeat; Signal-anchor; Sulfation;

KW Transmembrane.

KW INIT_MET 0 0

FT DOMAIN 1 12 Cytoplasmic (Potential).

FT TRANSMEM 13 33 Signal-anchor for type II membrane

FT protein (Potential).

FT DOMAIN 34 1856 Luminal (Potential).

FT DOMAIN 88 132 P-type 1.

FT DOMAIN 197 914 Maltase.

FT DOMAIN 952 998 P-type 2.

FT DOMAIN 1066 1812 Glucoamylase.

FT DOMAIN 37 83 Ser/Thr-rich.

FT ACT_SITE 528 528 Nucleophile (By similarity).

FT ACT_SITE 531 531 By similarity.

FT ACT_SITE 1419 1419 Nucleophile (By similarity).

FT ACT_SITE 1422 1422 By similarity.

FT ACT_SITE 1525 1525 Proton donor (By similarity).

FT DISULFID 89 117 By similarity.

FT DISULFID 100 116 By similarity.

FT DISULFID 111 129 By similarity.

FT DISULFID 965 982 By similarity.

FT DISULFID 977 995 By similarity.

FT MOD_RES 415 415 Sulfotyrosine (Potential).

FT	MOD_RES	424	424	Sulfotyrosine (Potential).
FT	MOD_RES	1281	1281	Sulfotyrosine (Potential).
FT	CARBOHYD	134	134	N-linked (GlcNAc. .) (Potential).
FT	CARBOHYD	294	294	N-linked (GlcNAc. .) (Potential).
FT	CARBOHYD	456	456	N-linked (GlcNAc. .) (Potential).
FT	CARBOHYD	457	457	N-linked (GlcNAc. .) (Potential).
FT	CARBOHYD	478	478	N-linked (GlcNAc. .) (Potential).
FT	CARBOHYD	706	706	N-linked (GlcNAc. .) (Potential).
FT	CARBOHYD	748	748	N-linked (GlcNAc. .) (Potential).
FT	CARBOHYD	826	826	N-linked (GlcNAc. .) (Potential).
FT	CARBOHYD	884	884	N-linked (GlcNAc. .) (Potential).
FT	CARBOHYD	911	911	N-linked (GlcNAc. .) (Potential).
FT	CARBOHYD	976	976	N-linked (GlcNAc. .) (Potential).
FT	CARBOHYD	988	988	N-linked (GlcNAc. .) (Potential).
FT	CARBOHYD	1254	1254	N-linked (GlcNAc. .) (Potential).
FT	CARBOHYD	1322	1322	N-linked (GlcNAc. .) (Potential).
FT	CARBOHYD	1363	1363	N-linked (GlcNAc. .) (Potential).
FT	CARBOHYD	1387	1387	N-linked (GlcNAc. .) (Potential).
FT	CARBOHYD	1602	1602	N-linked (GlcNAc. .) (Potential).
FT	CARBOHYD	1671	1671	N-linked (GlcNAc. .) (Potential).
FT	CARBOHYD	1841	1841	N-linked (GlcNAc. .) (Potential).
FT	CARBOHYD	1846	1846	N-linked (GlcNAc. .) (Potential).
SQ	SEQUENCE	1856	1856	AA; 209721 MW; 81E7AA0CABABD07D CRC64;

Query Match 29.2%; Score 1904.5; DB 1; Length 1856;

Best Local Similarity 43.8%; Pred. No. 1.7e-100;

Matches 395; Conservative 133; Mismatches 327; Indels 47; Gaps 21;

QY	348	PCRPRAPVTCQDVP---PNSRFDCAFDKAITQCEARGCCYIPAKQKQGLQGAQMPQWCF	404
DB	77	PQTGTTPVSACPVNLELRINCIPDQPTKATCDQRCCKNP-----QGA-VSPWCY	130
QY	405	FPSPSYSKLE-NLSSEMGYATLTRTPT--FPKDLTLRLDVMETENLRHFTIKD	461
DB	131	YSKNH-SYHVEGNLVNTNAGFTARL-KNLPSPVFGSNVDNVLTAEYQTSNRPFKLT	188
QY	462	PANRYEVEPLETRVHS--RAPSPLYSVFSEPEPGVIVHRQLDGRVLNTTVAPIFFAD	519
DB	189	QTNRFEPVPHVQSFSGNAAASLTQVVEISQPFISIKVTRRSNNRNFDDSGPLLFAD	248
QY	520	QFLQLSTSLPSQVITGLAEHL-SPLMLSTSWTRITLWNRDLAPT-PCANLYGSHPFYAL	577
DB	249	QFLQLSTRLPSNTVYGLGHEVHQVHRDNNWKTWTFINRDTTPNGNGTLYGAQTFFCL	308
QY	578	ED-GGSAHGCVLLNSNAMDVLQSPALSWRSTGGILDVYIFLGPPEKSVVQOYLDVGY	636
DB	309	EDASGLSGCVFLMNSNAMEVVLQAPAITRTIGTIGILDVYFVLGNTPEVQVQYELIGR	368
QY	637	PMPPYWGGLGFLCRWGYSTAITQVENVMTTRAHPFLDQVNDLDYMSRDRFTFNKDG	696
DB	369	PALPSYWGGLGFLSRYYEYGTLDNMRREVVERNRAAQLPYDVQHADIDYMDERRDFTYDSD	428
QY	697	FRDFPAMVQELHGGRRYMMIVDPAIS-SSGPAGSVRPYDEGLRRGVFTTNETG-OPLIG	754
DB	429	FKGFFPVELNHNQOKLVIIIDPAISNNSSSKPYGPIYDRGSDMKIWNSSDGVTPGIC	488
QY	755	KYWPGSTAPPDFTNPTALAWEDMVAEFDQVDFDGLWIDMNEPSNFIKRGSEDCPNNEL	814
DB	489	EWMPGTVPDPTNCPNCAVWTKFELFHQVDFDGLWIDMNEVSNFVDSVSGCSTNNL	548
QY	815	ENPPVPGVVGTLQQAATTCASHQFLSTHYLNHLXGLTEALASHRALYKA-RGTRPFV	873
DB	549	NNPPTPRILDGLYFCKTLCMDAVQHGWKQYDIHNLGYGSMATAEAATAKVPNKRSTI	608
QY	874	ISRSTFAGHRYAGHWTGDMSSWEQLASSVPEILQFNLGLVPLGADVCGFLGNTSEEL	933
DB	609	LTRSTFAGSGKFAAHLWGNTATWDLRWSIPGVLEFNLFGIPMWGPDICGFALDTPPEL	668
QY	934	CYRWTQLGAFYFPFMNHNLSLSPQEPYSFSEPA--QQAMRKALTLRYALLPHLYLPHQ	991
DB	669	CRWNLGAFYFPRNHNQGYKDDQDPASFGADSLNLSNRHLYNIRYTLFPVLYLTFPR	728
QY	992	AHVAGETVARPLFLBFFPKDSSSTWTVVDHQLLWGEALLITPVLAGKAEVGTGYPLGTWYDL	1051

Db 729 AHSRGDTVARPLLEHFEYEDNSTWDHQQFLWGPGLLITPVLDEGAEKVMAYVPDVAWYDY 788
Qy 1052 QVPIEALGSLPPPAAPREPAHSEGOVWTLPAPLDTINVLHAGYIIPLOGPGLTTE 1111
Db 789 ET-----GS-----QVRWKQKQVEMELFODKIGLHRLGYPFTQOPNTTTLA 831
Qy 1112 SRQCPMALAVALTGGGARGELFWDGSLVLERGAYTQVIF-LARNNTIYNELVRVTS 1170
Db 832 SRKNPLGLIILDENKEAKGELFWDGGETKDTVANKVYLCEFSVTQNRLEVNISQSTVK 891
Qy 1171 EGAGLOLQKVTVLGVATAPQ-QVLSNGVPVS---NFTYSPDTKVLIDIC-VSLIMGEQFLV 1225
Db 892 DPNNAFNEIKILGTBEPNSVTVKNGVPSQTSPTVTYDSNLKVAITIDILLGEAYTV 951
Qy 1226 SW 1227
Db 952 EW 953

RESULT 9
SUIS_HUMAN
ID SUIS_HUMAN STANDARD; PRT; 1826 AA.
AC P14410;
DT 01-JAN-1990 (Rel. 13, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Sucrase-isomaltase, intestinal [Contains: Sucrase (EC 3.2.1.48);
DE Isomaltase (EC 3.2.1.10)].
GN Name=SI;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RX MEDLINE=92359963; PubMed=1353958;
RA Chantret I., Lacasa M., Chevalier G., Ruf J., Islam I., Mantel N.,
RA Edwards Y., Swallow D., Rousset M.;
RT "Sequence of the complete cDNA and the 5' structure of the human
RT sucrase-isomaltase gene. Possible homology with a yeast
RT glucoamylase.";
RL Biochem. J. 285:915-923 (1992).
RN [2]
RP SEQUENCE OF 1-677 FROM N.A.
RX MEDLINE=88112852; PubMed=2962903; DOI=10.1016/0378-1119(87)90181-8;
RA Green F., Edwards Y., Hauri H.-P., Povey S., Ho M.W., Pinto M.,
RA Swallow D.;
RT "Isolation of a cDNA probe for a human jejunal brush-border hydrolase,
RT sucrase-isomaltase, and assignment of the gene locus to chromosome
RT 3.";
RL Gene 57:101-110 (1987).
RN [3]
RP SEQUENCE OF 1-19 AND 1007-1023, AND TISSUE SPECIFICITY.
RX PubMed=1677636;
RA Gervel J.P., Ferrero A., Chambrud L., Rigal A., Bonicel J.,
RA Maroux S.;
RT "Expression of sucrase-isomaltase and dipeptidylpeptidase IV in human
RT small intestine and colon.";
RL Gastroenterology 101:618-625 (1991).
RN [4]
RP VARIANT DISACCHARIDE INTOLERANCE I PRO-1097.
RX MEDLINE=96189940; PubMed=8609217;
RA Ouwendijk J., Moolenaar C.E.C., Peters W.J., Hollenberg C.P.,
RA Ginkel L.A., Franssen J.A.M., Naim H.Y.;
RT "Congenital sucrase-isomaltase deficiency: identification of a
RT glutamine to proline substitution that leads to a transport block of
RT sucrase-isomaltase in a pre-Golgi compartment.";
RL J. Clin. Invest. 97:633-641 (1996).
CC -1- FUNCTION: Plays an important role in the final stage of
CC carbohydrate digestion.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of sucrose and maltose by an alpha-

D-glucosidase-type action.
-1- CATALYTIC ACTIVITY: Hydrolysis of 1,6-alpha-D-glucosidic linkages
in some oligosaccharides produced from starch and glycogen by
alpha-amylase, and in isomaltose.
-1- SUBUNIT: The resulting sucrase and isomaltase subunits stay
associated with one another in a complex by non-covalent linkages.
-1- SUBCELLULAR LOCATION: Type II membrane protein. Brush border.
-1- TISSUE SPECIFICITY: Expressed in the poorly differentiated crypt
cells of the small intestine as well as in the mature villous
cells. Expressed at very low levels in the colon.
-1- PTM: The precursor is proteolytically cleaved when exposed to
pancreatic proteases in the intestinal lumen.
-1- PTM: Sulfated (By similarity).
-1- DISEASE: Defects in SI are the cause of disaccharide intolerance I
[MIM:222900].
-1- MISCELLANEOUS: There is a high degree of homology between the
isomaltase and sucrase portions (41% of amino acid identity)
indicating that this protein is evolved by partial gene
duplication.
-1- SIMILARITY: Belongs to the glycosyl hydrolase 31 family.
-1- SIMILARITY: Contains 1 P-type (trefoil) domain.
-1- SIMILARITY: Contains 1 P-type (trefoil) domain.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
EMBL; X63597; CAA45140.1; -
EMBL; M22616; AAA60551.1; ALT_SEQ.
PIR; S36082; UUUU.
Genew; HGNC:10856; SI.
MIM; 222900; -
GO; GO:0005903; C:brush border; TAS.
GO; GO:0005794; C:Golgi apparatus; TAS.
InterPro; IPR000322; Glyco_hydro_31.
InterPro; IPR000519; P_trefoil.
Pfam; PF01055; Glyco_hydro_31; 2.
Pfam; PF00088; Trefoil; 2.
SMART; SM00018; PD; 2.
PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 2.
PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.
PROSITE; PS00025; P_TREFOIL; 1.
Direct protein sequencing; Disease mutation; Glycoprotein;
Glycosidase; Hydrolase; Multifunctional enzyme; Repeat; Signal-anchor;
Sulfation; Transmembrane.
INIT_MET 0
CHAIN 1 1826 Sucrase-isomaltase, intestinal.
CHAIN 1 1006 Isomaltase.
CHAIN 1007 1826 Sucrase.
DOMAIN 1 11 Cytoplasmic.
TRANSMEM 12 31 Signal-anchor for type II membrane
protein.
DOMAIN 32 1826 Lumenal.
DOMAIN 61 108 P-type.
DOMAIN 109 1006 Isomaltase.
DOMAIN 1007 1826 Sucrase.
DOMAIN 42 59 Ser/Thr-rich.
ACT_SITE 504 504 Nucleophile.
ACT_SITE 507 507 By similarity.
ACT_SITE 1393 1393 Nucleophile.
ACT_SITE 1396 1396 By similarity.
ACT_SITE 1499 1499 Proton donor (By similarity).
DISULFID 62 93 By similarity.
DISULFID 76 92 By similarity.
DISULFID 87 105 By similarity.
MOD_RES 236 236 Sulfotyrosine (potential).
MOD_RES 238 238 Sulfotyrosine (potential).
MOD_RES 390 390 Sulfotyrosine (potential).
MOD_RES 399 399 Sulfotyrosine (potential).
MOD_RES 666 666 Sulfotyrosine (potential).

```
FT MOD_RES 762 762 Sulfoxyrosine (Potential).
FT MOD_RES 764 764 Sulfoxyrosine (Potential).
FT CARBOHYD 98 98 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 436 436 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 454 454 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 822 822 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 854 854 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 903 903 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 925 925 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 1234 1234 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 1302 1302 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 1339 1339 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 1353 1353 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 1402 1402 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 1534 1534 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 1571 1571 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 1674 1674 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 1747 1747 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 1762 1762 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 1814 1814 N-linked (GlcNAc. .) (Potential).
FT VARIANT 1097 1097 Q -> P (in disaccharide intolerance I;
exhibits intracellular accumulation of
mannose-rich SI in the Golgi).
/FTId=VAR_007854.
V -> F (in Ref. 3).
Missing (in Ref. 2).
S -> E (in Ref. 3).
V -> T (in Ref. 3).
SEQUENCE 1826 AA; 209272 MW; 3P7E4B66FDCP9C8E CRC64;

Query Match 28.8%; Score 1877.5; DB 1; Length 1826;
Best Local Similarity 41.7%; Pred. No. 5.9e-99;
Matches 382; Conservative 157; Mismatches 317; Indels 59; Gaps 19;

QY 340 AEAECAHGRPRAVTQCDVPSNPFDCAPDKAIXTQCEARGCCYIPAKQGLQAGMG 399
DB 48 ATTRVTNDSKCNVLNDVNVNINCPIDPEQFTGICAQGCCWRPNWDSL----- 101
QY 400 QMCFPPSPVSKLENLSSEMGYATLTR-TTPTFFPKDILTLRDVNMETENLHFT 458
DB 102 IPKCFVDNH-GYVQDMTTSIGVEAKLRIPSPTLFGNDINSVLFTQNTQPNFRFK 160
QY 459 IKDANRRYEVPLETRVHS--RAPSLYSVEFBEPFVYHQRQDGRVLLMTVAPLF 516
DB 161 ITDPNNRRYEVPHQYKEFTGPTVSDLYDKVAQNPFISQVIRKSGKGLFTDSIGPLV 220
QY 517 FADQFLQSLSPQVITGLAEHL-SPLMLSTSWTRITLWNRDLAP-TPGANLYGSHPFY 574
DB 221 YSDQYLQISARLPSDIYIGEGQVHKRFRHDLWSKWTPIPTRDQLPGDNNNLYGHQTF 280
QY 575 LALED-GGSAHGFLNNSAMDVLPSPALSWRSTGGILDVYIFLGPEPKSVVQQYLDV 633
DB 281 MCIEDTSKSGFGVFLNNSAMETIQTPVITRVTVGGILDVFIILLGDTPEQVVQYQQL 340
QY 634 VGYPFMPYWGGLFHLRCMGYSSTAITRQVENMTRAHFFLDVQWDLDTMDSRRDFTN 693
DB 341 VGLPAMPAYNWLGFQLSRWNYKSLDVVKEVVRNREAGIPFDVQVTDIDVMDKQFTVD 400
QY 694 KDGFRDPPAMQELHOGGRYMMVDPATSSSGPAG--SYRPYDEGLRGVFTNQTGQ- 750
DB 401 QVAFNGLPQVQDLHDHQKGVILDPATISIGRRANGTYYATYERGNTQHWLINESDGST 460
QY 751 PLIGKVPKGTAPDFTNPTALAWEDMVAEFHDQVDFGLWMDMNEPSNFIKSGSDGCP 810
DB 461 PIIGVWPGLUTVYDFTNPCINDWANECSIFHQEVOYDGLWMDNEVSPFIOGSKYGCN 520
QY 811 NNELENPYPVGVGGTQAAITCASHQPLSTHYNLNLGLTEIASHARLVKA-RGT 869
DB 521 VNKLYNPPTFDILDKLMSKTIICMDAVQWQGVQDVHSLYGYSMATATQAVQKVPFNK 580
QY 870 RPFVISRSTAGHVRVAGHTGVDWNSWEQLASVPEILQFNLLGVPLVGADVCGFLGNT 929
DB 581 RSPILTRSTPAGSGRAAAHLWGNTASWEQMEWSITGMLEFSLFGIPLVGADICGFVAET 640
```

```
QY 930 SBEELVVRWTQLCGAFYFPFMNHNLSLSLPQEPYSFSEPA-----QQAMRKALTURYAL 981
DB 641 TBEELCRMMQLGAFYFPFSRNHNS-----DGYEHQDPAFFQGNLSLVKSSRQVLTIRYTL 694
QY 982 LPHLYTLFHOAHVAGETVARPLPLEFPKDSSTWTVDDHQLLWGEALLITPVLQAGKAEVIG 1041
DB 695 LPFLYTLFYKAHVGETVARPVLHFEFYDTSNWIETEFELMGPALDITPVLKQAGDVSA 754
QY 1042 YPLGWTYDLOTVPIEALGSLPAPPAAPREPAIHSEGOVVTLPAPLDTINVHLRAGYIIP 1101
DB 755 YIPDAIWDYES-----GAKRP-----WRKQVDMYLPADKIGLHRLGGYIIP 797
QY 1102 LQPGGLTTTTSROQPMALAVALTKGGEARGELFWDDGESLEVLERGAYTQVIFLARNTI 1161
DB 798 IQEPDVTTASRKNPLGLVALGENTAKGDFDDGETKDTIQNGYNYLYTFSVSNNTL 857
QY 1162 VNELRVVTS--EGAGLQLOKTVTVLGV--ATAPOQVLSNGVPV--SNFTYSPTDKVLDI 1213
DB 858 --DIVCTHSSSYQEGTTLAFQTVKILGLTSDVTEVVAENQPMNAHSNFTYDASNQVLLI 915
QY 1214 C-VSLIMGEOQLVSW 1227
DB 916 ADLKNLGNFNSVQW 930

RESULT 10
SUIS_RABIT STANDARD; PRT; 1826 AA.
AC P07768;
DT 01-AUG-1988 (Rel. 08, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Sucrase-isomaltase, intestinal [Contains: Sucrase (EC 3.2.1.48);
DE Isomaltase (EC 3.2.1.10)].
GN NamesSI;
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]_
RP SEQUENCE FROM N.A.
RX MEDLINE=86245068; PubMed=3755079; DOI=10.1016/0092-8674(86)90739-7;
RA Hunziker W., Spiess M., Semenza G., Lodish H.F.;
RT "The sucrase-isomaltase complex: primary structure, membrane-
RT orientation, and evolution of a stalked, intrinsic brush border
RT protein.";
RL Cell 46:227-234 (1986).
RN [2]
RP PRELIMINARY SEQUENCE OF 1-37 AND 1007-1014.
RX MEDLINE=83105704; PubMed=7152027; DOI=10.1016/0014-5793(82)80833-8;
RA Sjoestrom H., Noren O., Christiansen L.A., Wacker H., Spiess M.,
RA Bigler-Meier B., Rickli E.E., Semenza G.;
RT "N-terminal sequences of pig intestinal sucrase-isomaltase and pro-
RT sucrase-isomaltase. Implications for the biosynthesis and membrane
RT insertion of pro-sucrase-isomaltase.";
RL FEBS Lett. 148:321-325 (1982).
CC -!- FUNCTION: Plays an important role in the final stage of
CC carbohydrate digestion.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of sucrose and maltose by an alpha-
CC D-glucosidase-type action.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,6-alpha-D-glucosidic linkages
CC in some oligosaccharides produced from starch and glycogen by
CC alpha-amylase, and in isomaltose.
CC -!- SUBUNIT: The resulting sucrase and isomaltase subunits stay
CC associated with one another in a complex by non-covalent linkages.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Brush border.
CC -!- PTM: The precursor is proteolytically cleaved when exposed to
CC pancreatic proteases in the intestinal lumen.
CC -!- PTM: N- and O-glycosylated.
CC -!- PTM: Sulfated (By similarity).
CC -!- MISCELLANEOUS: There is a high degree of homology between the
CC isomaltase and sucrase portions (41% of amino acid identity)
```

indicating that this protein is evolved by partial gene duplication.

-I- SIMILARITY: Belongs to the glycosyl hydrolase 31 family.

-I- SIMILARITY: Contains 1 p-type (trefoil) domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M14046; AAA31459.1; -

DR InterPro; IPR000322; Glyco_hydro_31.

DR InterPro; IPR000519; p_trefoil.

DR Pfam; PF01055; Glyco_hydro_31; 2.

DR Pfam; PF00088; Trefoil; 2.

DR SMART; SM00018; PD; 2.

DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 2.

DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 2.

DR PROSITE; PS00025; P_TREFOIL; 1.

KW Direct protein sequencing; Glycoprotein; Glycosidase; Hydrolase;

KW Multifunctional enzyme; Repeat; Signal-anchor; Sulfation;

KW Transmembrane.

INIT MET 0

CHAIN 1 1826

CHAIN 1 1006

CHAIN 1007 1826

DOMAIN 1 11

TRANSMEM 12 31

DOMAIN 32 1826

DOMAIN 42 59

DOMAIN 61 108

DOMAIN 109 1006

DOMAIN 1007 1826

ACT SITE 504 507

ACT SITE 507 504

ACT SITE 1393 1393

ACT SITE 1396 1396

ACT SITE 1499 1499

DISULFID 62 93

DISULFID 76 92

DISULFID 87 105

MOD RES 390 390

MOD RES 399 399

MOD RES 1381 1381

MOD RES 1384 1384

CARBOHYD 41 41

CARBOHYD 98 98

CARBOHYD 454 454

CARBOHYD 858 858

CARBOHYD 895 895

CARBOHYD 903 903

CARBOHYD 1001 1001

CARBOHYD 1234 1234

CARBOHYD 1302 1302

CARBOHYD 1324 1324

CARBOHYD 1339 1339

CARBOHYD 1353 1353

CARBOHYD 1367 1367

CARBOHYD 1402 1402

CARBOHYD 1534 1534

CARBOHYD 1571 1571

CARBOHYD 1747 1747

CARBOHYD 1762 1762

CARBOHYD 1798 1798

SEQUENCE 1826 AA; 210008 MW; 6840D03955A45BB5 CRC64;

Query Match

Best Local Similarity 28.8%; Score 1875.5; DB 1; Length 1826;

Matches 377; Conservative 155; Mismatches 312; Indels 53; Gaps 19;

QY	355	PTQCDVPPNSRDCAPDKAITQEQCEARGCCYIPAKQGLQAGMGPQWCFPPSPSYKL	414
DB	63	PSSELNVNERINCIPESQPTQAIQAQNCWCWRPNNS-----DIPWCFVDNH-GYNV	115
QY	415	ENLSSENGYATLTR-TTPFPFKDILTLRLDVMMETENRLHFTIKDPANRYEVLB-	472
DB	116	EGMTTSTGLEARLNKSTPTLFGNDINNVLTTESQTANRLRFLKLTDPNNKRYEVP	175
QY	473	-TPRVHSRAPSLYSVESEPPGVIVHRLQDGRVLLNTTAPLFFAQLFOLSLSPSQ	531
DB	176	VTEFAGPATETLYDVQVTENPFISKIRKSNRILFDSISGLPLVYSOQLSTRPSE	235
QY	532	YITGLAHL-SPLMLSTSWTRITLWNRDL-APTGANLYGSHPPFYLALED-CGS	588
DB	236	YMYGFGEVHKFRHDLWKTPITFRQHTDDNNNLYGHQTFEMCIEDTTGKSGVFL	295
QY	589	LNSNADVVLPSPALSWRSTGGILDVIFIGPEPKSVVQVLDVVGVYFPMPPYGLG	648
DB	296	MNSNAMEIFIOPTPIVTVRVIIGILDVIFILGDTPEQVVQVQELIGRPAMPAY	355
QY	649	LCRWGYSSTAITROVVENMTRAHPFLDVQWDLVDMSRRDFTENKDGFRDP	708
DB	356	LSRWYNSLDVVVKEVRRNRREALIPFTQVSDIDMEDKDKFTYDRVANGLP	415
QY	709	QGGREYMMIVDPAIS-----SSGPAGSYRYPDEGLRRGVFITNETG-OPL	763
DB	416	DHGQKYVILDPAISINRRASGEA--YESYDRGNAQNVWVESDGTTPIVGEV	473
QY	764	PDFTNPTLAWEDMVAEFHQVDPDGLWIDMNEPSNFIKRGSDGCPNNELEN	823
DB	474	PDFTSPNCIEWANEENIFHOEVNYDGLWIDMNEVSVFQSGNKGCDNTLNP	533
QY	824	VGTTLOAATICASSHQFLSTHYNLNLYLGLTEATASHRALVKA-RGTSPF	882
DB	534	VDKLMYSKTLCDMSVQYWGKQYDVHSLGYSNAIATERAVERFPNKRSL	593
QY	883	GRYAGHWGDMVSSWEQLASSVPEILOFNLLGVPLVGADVCGFLGNTSEEL	942
DB	594	GRHAAHLGDNTATWEQWESITGMLEFGLFGMPLVGADICGFLAETTEL	653
QY	943	FYPFRNHNLSLSPQEPYSFSEPA--QOAMRKALTLYALLPHLYTYLHQ	1000
DB	654	FYPFRNHNADGFHQDPAFFGQDLSLVKSSRHYLNIRYTLPLLYTLFY	713
QY	1001	RPLFLFEPKDSSTVTDHQLLWGEALLITPVLQACKAEVTGYPGLGTW	1060
DB	714	RPVLHEFYEDTNSWVEDREFLWGPALLITPVLVTQAEVTSAYIPDAV	767
QY	1061	SLPPPPAAPREPAIHSEQGWTLPAPLDTINVHVRAGYIIPQGLTTSRQ	1120
DB	768	AKRP-----WRKQVEMSLPADKIGLHAGGYIIPQPAVTTASRMN	816
QY	1121	VALTKGBARGELFWDDGESLEVBERGAYTVQVIFLARNN-----	1176
DB	817	TALDNDTAVGDFFWDDGCTKDTQNDVNYLYTFVAVSNNNLNITCTHEL	873
QY	1177	LQKVTVLGVATAPQ--VLSNGVPV--SNFTYSPDKVLDI-CVSLMGE	1227
DB	874	FQTIKILGVTETVQTVVAENNSQMSHNSFTYSPDKVLDI-CVSLMGE	930
RESULT 11			
Q8TE24			
ID	Q8TE24	PRELIMINARY; PRT; 1734 AA.	
AC	Q8TE24;		
DT	01-JUN-2002 (TrEMBLrel. 21, Created)		
DT	01-JUN-2002 (TrEMBLrel. 21, Last sequence update)		
DT	01-JUN-2003 (TrEMBLrel. 24, Last annotation update)		
DE	Maltase-glucoamylase (Fragment).		
GN	Name=MGAM;		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN NCBI_TaxID=9606;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22457205; PubMed=12547908; DOI=10.1073/pnas.02371170100;
 RA Nichols B.L., Avery S., Sen P., Swallow D.M., Hahn D., Sterchi E.;
 RT "The maltase-glucoamylase gene: common ancestry to sucrase-isomaltase
 with complementary starch digestion activities.";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:1432-1437(2003).
 RN SEQUENCE FROM N.A.
 RP Nichols B.L. Jr., Avery S.E., Sen P., Swallow D.M., Hahn D.,
 RA Sterchi E.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF432202; AAL83560.1; JOINED.
 DR EMBL; AF432187; AAL83560.1; JOINED.
 DR EMBL; AF432188; AAL83560.1; JOINED.
 DR EMBL; AF432189; AAL83560.1; JOINED.
 DR EMBL; AF432190; AAL83560.1; JOINED.
 DR EMBL; AF432191; AAL83560.1; JOINED.
 DR EMBL; AF432192; AAL83560.1; JOINED.
 DR EMBL; AF432193; AAL83560.1; JOINED.
 DR EMBL; AF432194; AAL83560.1; JOINED.
 DR EMBL; AF432195; AAL83560.1; JOINED.
 DR EMBL; AF432196; AAL83560.1; JOINED.
 DR EMBL; AF432197; AAL83560.1; JOINED.
 DR EMBL; AF432198; AAL83560.1; JOINED.
 DR EMBL; AF432199; AAL83560.1; JOINED.
 DR EMBL; AF432200; AAL83560.1; JOINED.
 DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR InterPro; IPR000322; Glyco_hydro_31.
 DR InterPro; IPR000519; P_trefoil.
 DR Pfam; PF01055; Glyco_hydro_31; 2.
 DR Pfam; PF00088; Trefoil; 1.
 DR SMART; SM00018; PD; 1.
 DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 2.
 FT NON_TER 1
 SQ SEQUENCE 1734 AA; 196946 MW; BAF70D11BEDDBA7E CRC64;
 Query Match 28.1%; Score 1830; DB 2; Length 1734;
 Best Local Similarity 44.4%; Pred. No. 3e-96;
 Matches 376; Conservative 129; Mismatches 303; Indels 38; Gaps 18;
 QY 401 PNCFFPPSPYSKLE-NLSSSEMGYATITRTPT--FPFKDILTLRLDMMETENLRHF 457
 DB 5 PCYYSKNH-SYHVEGLNVTNAGFTARL-KNLPSPVFGSNVDNVLTAETSNRHF 62
 QY 458 TKDPAIRRYEVPLETPRVHS--RAPSPLYSEFFSEEPFGVIVHROLDGRVLLNTTVAPL 515
 DB 63 KLTQDTNRFEPVHEHVQSFSGNAAALTYQVEISRQPSIKVTRSNRNVLPDSSIGPL 122
 QY 516 PFADQFLQSTSLPSQYITGLAEHL-SPLMLSTSWTRITLWNRDLAPT-PGANLYGSHPF 573
 DB 123 LFADQFLQSTSLPSQYITGLAEHL-SPLMLSTSWTRITLWNRDLAPT-PGANLYGSHPF 182
 QY 574 YLALED-GGSAHGAVFLNNDVVLQSPALSSWRSTGGILDYVIFLGPESKSVVQOYLD 632
 DB 183 FLCLEDAAGLFGVFLNNDVVLQSPALSSWRSTGGILDYVIFLGPESKSVVQOYLD 242
 QY 633 VVGYPMPYPYGLGFLHLCRWGYSSTAITROQVENMTTRAHPDLDVQNDLDYDMSRRDFTF 692
 DB 243 LIGRPALPSYALGFLHLSRYEYGLTLDNMRVVERNRAQLPYDVQHADIDYMDERRDFTY 302
 QY 693 NKDGRDPFAMVQELHOGGRRYIMVYDPAIS-SSGPAQSYRYPVDEGLRGVFTNETG-Q 750
 DB 303 DSVDFKGFPEFVNEHLNNGOKLIIIVDPAISNNSSSKPYGPDGSDGVKVIWNSDGV 362
 QY 751 PLIGKVPKSTAPPDFTNTPTALAWEDMVAEFHQVFPDGLWIDMNEPNSFNIRGSDGCP 810
 DB 363 PLIGVWPQGVTPDPTNPNCAVWTKPELFFNQVEFDGIDMNEVENFVDSVSGCS 422
 QY 811 NNELENPPYVPGVGGTLOQAATTCASSHQFLSTHYNLHNLGLTEAIAASHRALVKA-RGT 869
 Db 423 TNNLANNPPPTPRILDGYLFCCKTLCMDAVQHWKQKQDIHNLGYSMATAAABAAKTVFPNK 482
 QY 870 RPFVISRSTFAGHGRYAGHWTGDVWSSWEQLASSVPEILQFNLLGVPLVGVADVCGFLGNT 929
 Db 483 RSPILTRSTFAGSGKFAHNLGDNTATWDDLWSPGVLEFNLFGIPMGVDPDCCGALDT 542
 QY 930 SBEICVRWTQLCAGYFPMRNHNSLLSLPQEPVSFSEPA--QOAMRKALTLRYALLPHLYT 987
 Db 543 PEELCRNMQLGAFYFPMRNHNSLLSLPQEPVSFSEPA--QOAMRKALTLRYALLPHLYT 602
 QY 988 LFHQAHVAGETVARPLFLFEPKDSSTWTVHQLLAGEALLITPVLQAGKAEVGYEPLGT 1047
 Db 603 LFFRAHSGDITVARPLLHFEYEDNSTWVHQFLMGPGGLITPVLDEGAKKMAVYVDAV 662
 QY 1048 WYDLQVTPIEALGSLPPPPAAPREPAIHSEGQWVTLPAPLDTINVHLRAGYIIPLOGPGL 1107
 Db 663 WYDYET-----GS-----QVRWRKQKVEMLPGDKLGLHLRGYIFPTQOPNT 705
 QY 1108 TTTERSQPMALAVALTGGEGARGELFWDGSGLEVLERGAYTVIFLARNNTIVNELVR 1167
 Db 706 TTLASRKNPLGLIIAALDENKEAKGELFWDGSGLEVLERGAYTVIFLARNNTIVNELVR 765
 QY 1168 VT-SEGAGLQKQVTLGVATAPQ-QVLSNGVPVS---NFTYSPDTKVLDDIC-VSLLMGE 1221
 Db 766 STYKDPNNLAFNEIKILGMBSPSNVTVXHGVPVSQTSPTVTYDSNKLVAITDIDLLGE 825
 QY 1222 QFLVSW 1227
 Db 826 AYTVEM 831
 RESULT 12
 SUIRAT STANDARD; PRT; 1840 AA.
 AC P23739;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 25-OCT-2004 (Rel. 45, Last annotation update)
 DE Sucrase-isomaltase, intestinal [Contains: Sucrase (EC 3.2.1.48);
 DE Isomaltase (EC 3.2.1.10)].
 GN Name-Si;
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Intestine;
 RX MEDLINE=95121929; PubMed=7821806; DOI=10.1016/0378-1119(94)90452-9;
 RA Chandrasena G., Osterholm D.E., Sunitha I., Henning S.J.;
 RT "Cloning and sequencing of a full-length rat sucrase-isomaltase-
 encoding cDNA";
 RL Gene 150:355-360(1994).
 RN [2]
 RP SEQUENCE OF 86-361 FROM N.A.
 RC STRAIN=Fischer 344; TISSUE=Intestine;
 RX MEDLINE=91097578; PubMed=2268340;
 RA Traher P.G.;
 RT "Regulation of sucrase-isomaltase gene expression along the crypt-
 villus axis of rat small intestine.";
 RL Biochem. Biophys. Res. Commun. 173:765-773(1990).
 RN [3]
 RP SEQUENCE OF 732-1372 FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Duodenum;
 RX MEDLINE=90381315; PubMed=2400788; DOI=10.1016/0167-4781(90)90121-H;
 RA Broyart J.-P., Hugot J.-P., Perret C., Porteu A.;
 RT "Molecular cloning and characterization of a rat intestinal sucrase-
 isomaltase cDNA. Regulation of sucrase-isomaltase gene expression by
 sucrose feeding.";
 RL Biochim. Biophys. Acta 1087:61-67(1990).
 RN [4]
 RP SEQUENCE OF N-TERMINUS OF ISOMALTASE AND SUCRASE.

RX	MEDLINE=82167542; PubMed=6802834;	FT	MOD_RES	1386	1386	Sulfotyrosine (Potential).
RA	Hauri H.-P., Wacker H., Rickli E.E., Bigler-Meier B., Quaroni A.,	FT	CARBOHYD	108	108	N-linked (GlcNAc. .) (Potential).
RA	Semenza G.;	FT	CARBOHYD	463	463	N-linked (GlcNAc. .) (Potential).
RT	"Biosynthesis of sucrose-isomaltase. Purification and NH2-terminal	FT	CARBOHYD	757	757	N-linked (GlcNAc. .) (Potential).
RT	amino acid sequence of the rat sucrose-isomaltase precursor (pro-	FT	CARBOHYD	764	764	N-linked (GlcNAc. .) (Potential).
RT	sucrose-isomaltase) from fetal intestinal transplants.";	FT	CARBOHYD	866	866	N-linked (GlcNAc. .) (Potential).
RL	J. Biol. Chem. 257:4522-4528(1982).	FT	CARBOHYD	909	909	N-linked (GlcNAc. .) (Potential).
CC	-I- FUNCTION: Plays an important role in the final stage of	FT	CARBOHYD	1239	1239	N-linked (GlcNAc. .) (Potential).
CC	carbohydrate digestion.	FT	CARBOHYD	1307	1307	N-linked (GlcNAc. .) (Potential).
CC	-I- CATALYTIC ACTIVITY: Hydrolysis of sucrose and maltose by an alpha-	FT	CARBOHYD	1344	1344	N-linked (GlcNAc. .) (Potential).
CC	D-glucosidase-type action.	FT	CARBOHYD	1358	1358	N-linked (GlcNAc. .) (Potential).
CC	-I- CATALYTIC ACTIVITY: Hydrolysis of 1,6-alpha-D-glucosidic linkages	FT	CARBOHYD	1372	1372	N-linked (GlcNAc. .) (Potential).
CC	in some oligosaccharides produced from starch and glycogen by	FT	CARBOHYD	1484	1484	N-linked (GlcNAc. .) (Potential).
CC	alpha-amylase, and in isomaltose.	FT	CARBOHYD	1512	1512	N-linked (GlcNAc. .) (Potential).
CC	-I- SUBUNIT: The resulting sucrose and isomaltase subunits stay	FT	CARBOHYD	1574	1574	N-linked (GlcNAc. .) (Potential).
CC	associated with one another in a complex by non-covalent linkages.	FT	CARBOHYD	1761	1761	N-linked (GlcNAc. .) (Potential).
CC	-I- SUBCELLULAR LOCATION: Type II membrane protein. Brush border.	FT	CARBOHYD	1828	1828	N-linked (GlcNAc. .) (Potential).
CC	-I- PTM: The precursor is proteolytically cleaved when exposed to	FT	CONFLICT	86	86	C -> W (in Ref. 1).
CC	pancreatic proteases in the intestinal lumen.	FT	CONFLICT	91	91	H -> S (in Ref. 2).
CC	-I- PTM: Sulfated (By similarity).	FT	CONFLICT	94	94	K -> Q (in Ref. 2).
CC	-I- MISCELLANEOUS: There is a high degree of homology between the	FT	CONFLICT	222	222	C -> F (in Ref. 2).
CC	isomaltase and sucrose portions (41% of amino acid identity)	FT	CONFLICT	230	230	L -> V (in Ref. 2).
CC	indicating that this protein is evolved by partial gene	FT	CONFLICT	240	240	T -> R (in Ref. 2).
CC	duplication.	FT	CONFLICT	252	252	G -> E (in Ref. 2).
CC	-I- SIMILARITY: Belongs to the glycosyl hydrolase 31 family.	FT	CONFLICT	283	283	L -> F (in Ref. 2).
CC	-I- SIMILARITY: Contains 1 p-type (trefoil) domain.	FT	CONFLICT	294	294	G -> E (in Ref. 2).
CC	-----	FT	CONFLICT	326	326	T -> A (in Ref. 2).
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	FT	CONFLICT	350	352	YWN -> FNG (in Ref. 2).
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	FT	CONFLICT	359	361	S -> A (in Ref. 3).
CC	the European Bioinformatics Institute. There are no restrictions on its	FT	CONFLICT	732	732	S -> A (in Ref. 3).
CC	use by non-profit institutions as long as its content is in no way	FT	CONFLICT	735	735	E -> V (in Ref. 3).
CC	modified and this statement is not removed. Usage by and for commercial	FT	CONFLICT	841	841	E -> Q (in Ref. 3).
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	FT	CONFLICT	915	915	A -> T (in Ref. 3).
CC	or send an email to license@isb-sib.ch).	FT	CONFLICT	923	923	A -> R (in Ref. 3).
CC	-----	FT	CONFLICT	929	930	AG -> GT (in Ref. 3).
DR	EMBL; L25926; AAA5097.1; -;	FT	CONFLICT	937	938	CR -> SQ (in Ref. 3).
DR	EMBL; M62889; AAA42144.1; -;	FT	CONFLICT	958	961	GTCT -> ETDK (in Ref. 3).
DR	EMBL; X15546; CAA33552.1; -;	FT	CONFLICT	979	979	Y -> C (in Ref. 3).
DR	PIR; T10799; T10799.	FT	CONFLICT	985	985	N -> H (in Ref. 3).
DR	RGD; 3675; Si.	FT	CONFLICT	996	997	LP -> SL (in Ref. 3).
DR	InterPro; IPR000322; Glyco_hydro_31.	FT	CONFLICT	1009	1009	P -> A (in Ref. 3).
DR	InterPro; IPR000519; P_trefoil.	FT	CONFLICT	1022	1022	T -> P (in Ref. 3).
DR	Pfam; PF01055; Glyco_hydro_31; 2.	FT	CONFLICT	1026	1026	G -> E (in Ref. 3).
DR	Pfam; PF00088; Trefoil; 1.	FT	CONFLICT	1031	1031	P -> K (in Ref. 3).
DR	SMART; SM00018; PD; 2.	FT	CONFLICT	1093	1093	R -> S (in Ref. 3).
DR	PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 2.	FT	CONFLICT	1098	1098	G -> A (in Ref. 3).
DR	PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; FALSE_NEG.	FT	CONFLICT	1301	1301	A -> D (in Ref. 3).
DR	PROSITE; PS00025; P_TREFOIL; 1.	FT	CONFLICT	1336	1336	P -> A (in Ref. 3).
KW	Direct protein sequencing; Glycoprotein; Glycosidase; Hydrolase;	FT	CONFLICT	1338	1339	VW -> WG (in Ref. 3).
KW	Multifunctional enzyme; Repeat; Signal-anchor; Sulfation;	SQ	SEQUENCE	1840	1840	AA; 210218 MW; 079ADA45E9A23E2E CRC64;
KW	Transmembrane.					
FT	INIT_MET 0					Query Match
FT	CHAIN 1					Best Local Similarity
FT	CHAIN 1013					27.8%; Score 1813.5; DB 1; Length 1840;
FT	TRANSMEM 12 31					41.8%; Pred. No. 2.9e-95;
FT	DOMAIN 32 1840					Matches 378; Conservative 146; Mismatches 318; Indels 63; Gaps 20;
FT	DOMAIN 71 118					
FT	DOMAIN 119 1012					
FT	DOMAIN 1013 1840					
FT	DOMAIN 42 69					
FT	ACT_SITE 513 513					
FT	ACT_SITE 516 516					
FT	ACT_SITE 1398 1398					
FT	ACT_SITE 1401 1401					
FT	ACT_SITE 1511 1511					
FT	DISULFID 72 103					
FT	DISULFID 86 102					
FT	DISULFID 97 115					
FT	MOD_RES 400 400					
FT	MOD_RES 409 409					
QY	351 PRAVPTQCDVPNPSRFDCAPKAITQEQCEARGCCYIPAKOGLQAGMQGPPFPSPYP 410					
DB	69 PGKCPPEQGEPINERINCIPQHPHTKAICEERGCGCWPNWNTV-----IPWCFADNH- 121					
QY	411 SYKLENLSSEMGYTATLTR-TTPTFFPKDILTLDVMMETENRHLFTIKDPANRREYV 469					
DB	122 GYNAESITNENAGLKATLNRIPTLPFGEDIKSVILTTQTGTGNFRFKITDPNNKREYV 181					
QY	470 PLETRVHSRAPS--PLYSVFSFSEPPFGVIVHRQLDGRVLLNTTVAPLFFAQDFQLQST 527					
DB	182 PHQFVKETGTPAADTLVDVQVSENPFSIKVIRKSNKVLCDTSVGPLLYSNQYLQISFR 241					
QY	528 LPSQVITGLAHL-SPLMLSTSWTITLWNRDLAP-TPCANLYGSHPFYLALED-CGSAH 584					
DB	242 LPSEVIYFGGHIHKRRFRHDLYKWTWPIFTFTRDEIPGDNHNLHYGHTFFMGIGDTSKSY 301					
QY	585 GVFLNSNAMDVLPSPALSRSRSTGGILDVYIFLGPEPKSVVQOYLDVGVGPPFPYWG 644					

SQ	SEQUENCE	1812 AA; 208172 MW; 3ED93E407D783158 CRC64;	AMRP_HUMAN	STANDARD;	PRT; 357 AA.
Query Match	27.7%; Score 1806.5; DB 1; Length 1812;				
Best Local Similarity	40.4%; Pred. No. 71e-95;				
Matches	365; Conservative 163; Mismatches 315; Indels 61; Gaps 21;				
QY	354 VPTQCDVPNSRPD----CAPDKAITQEOCEARGCCYIPAKOGLQAGQWQCWCFPPPSY 409				
DB	43 IPGKCPAENRDLDEKINCIPDQFPQALCAMGCCWCPNPNES-----PTWFCSPANNH 96				
QY	410 PSYKLENLASSEMGYATITRT-PTFFPKDILTLVDWMTENRLHTIIPANRVE 468				
DB	97 -GYEFKISNPINFENLKNKSPPTFGDNIITLLTTQSTANRFRKFTDPNNQRYE 155				
QY	469 VPLETPRVHSRAP--SPLYSVEFSEPFQVIVHROLDGRVLLTNTVAPLPADQFLQS 525				
DB	156 VHQFVKNKDFGPPASNPIDYKITEPISIKVIRASNNKIILFDSIGPLVTSNQVLQIS 215				
QY	526 TSLPSQYITGLAEHL-SPLMLSTSMTITLWNRDLAP-TPGANLYGSHPHYLALED-GGS 582				
DB	216 TKLPSKYIYGLGEVHKRFRHLDYKWTPIFTRDQLPGDNNNNLYGHQTFNMSIEDTSCK 275				
QY	583 ANGCVLLNSAMDVLOPSALSWRSTGGILDVYIFELGPEPKSVVOQYLDVGVYFPMPPY 642				
DB	276 SFGVFLMNSNAMEVFIQPTPIVTVIGILDYIFELGDTPGQVQOQYQELTGRPAMPY 335				
QY	643 WGLGFHLCKWGSYSTAITRQVVENMTRAHFFLDVQWNLDDYMSRRDFTFNKDGFRDFA 702				
DB	336 WSLGFQLSRWYGSLDVAVKVENRDRARIPDAQVTDIDYMEKDKFTYNNKTFYGLPE 395				
QY	703 MYQELHQGRRYMMIVDPAISSGPGS--YRPYDEGLRRGVFITNETG-QPLIGKWPFG 759				
DB	396 FVKDLHDHQKVIILDPAISITSLANGNHYKTYERQNEQKVWYQSDGTTLGIEVWPG 455				
QY	760 STAFDPFTPTALAWEDVAFHDQVDPDGLWIDWEPNSFIRSGDCPNNELENPPY 819				
DB	456 LTVYDPFTNPKLDWMTNCSIFHEIKYDGLWIDMNEVSSRFVHSGTSCSKNKLNYPPF 515				
QY	820 VPGVGGTLOAATICASSHQFLSTHNLNLYGLTEAIAASHALVKA-RGTGPFVTSRT 878				
DB	516 IPDILDKMYAKTICMDAIQHWGKQYDVHSLYGYSAIATEKAEKVPFNKSFILTRST 575				
QY	879 FAGHGYAGHTGDVMSWEQLASSVPEILQFNLLGVPLVGDVCGFLGNTSEELCVRWT 938				
DB	576 FAGTGKHATHWLGDNTPSEHMEWSITPMLFGLGMPFAGIDICGVVVDTEELCRWM 635				
QY	939 QLGAFYFPRNHNLSLPOEYSEFA--QOMRKALTLRYALLPHLYTLFHOAHVAG 996				
DB	636 QIGAFYFPRDNAGGYMPQDPAYFGQDSLNVNTRSHYLDIWTLLPYLYNLLYKAYVYG 695				
QY	997 ETVARPLFEFPKDSSTWVDHQLWGEALLITPVLOAGKAEVTGYFPLGTWYDLQTVPI 1056				
DB	696 ETVARPLFEFFEDTNSWIEDQLWGSALLITPVLROGADMSAYIPDATWYDET--- 752				
QY	1057 EALGSLPPPPAAPREPAIHSEGW----VTLPAPLDTINVHLRAGYIIPLOGPGLTTTES 1112				
DB	753 -----GKRTWRQRVEMVLPBGKIGLHVRGGYIIPQCPAVNTTAS 794				
QY	1113 RQPMALVALYTKGEARGELFWDDGESLEVERGAYTOVIFLARNTIVNELVRVTS--- 1170				
DB	795 RKNPLGLIILADNNA-AGKDFFWDDGESKDEKGYILYTFVSLNNEL--DIICHTSSY 851				
QY	1171 -EGAGLQLOKVTVLGVA-TAPO-QVLSNG---VPVSNFTYSPDTKVLDI-CVSLMGEQF 1223				
DB	852 QEGTTLAFTIKILGLANTVTQVAVENNQOIIHNSFTYHASNQSLIIDNLKLNKGNF 911				
QY	1224 LVSW 1227				
DB	912 TVQW 915				

RESULT 14

AMRP_HUMAN

Db 299 GYKQVSDIITNVVENYNKSIQPLDVIWDDDDHMDGAKDFTLDPINYPEYKLRPPFLDRIHAN 358
Qy 711 GRZYMIIDPAISSSGPAGSYRDEGLRGVFIITNETQOPLIGKVMRGSTAFPPDTNET 770
Db 359 GMYVVLIIIDPGIAIN---TSYGTFRQGMADDVFI-KHGGSPFLGQWPGAVYFPDFLNP 414
Qy 771 ALAWEDMVAEFHQVDFDGLWIDMNEPSNFIRG-----SEDGCPN----- 811
Db 415 TVNFWADEISHFHQWVPDGLWIDMNEISNFCSCGKSIPTNRSCTGCTGPFWECCLDRTNI 474
Qy 812 --NELNPPYVPGVGG--TLQAATTCASHQFLSTHYN-----LHNYGLTEAIAASHR 861
Db 475 TATRWDPVPYKINASCTQVPLGFKTIATS-----SVHYNGVLEYDAHSLYGLSQAIAATHK 529
Qy 862 ALVKARGTRPFIISRSTFAGHGYAGHWTGCDVWSSWEQLASSVPEILOFNLLGVPLVGAD 921
Db 530 ALQNLLDKRPFVLTSTFVSGSYAAHWTGDNKATWEDLRYSTITLNFMGFMGMVVGAD 589
Qy 922 VCGFLGNTSEELCVRWTLQCAFYPFMRNHNLSLFPQEPYFSEPAQQAAMRKALTRLRYAL 981
Db 590 ICGFYDITTEELGCRWILQCAFYFSDHNSLASKQELYLWDSVAKSA-RKALGLGYEL 648
Qy 982 LPHLYTLFHQAHVAGETVARPLEPPKDSSTWTVDHLLMGEALLITPVLOAGKAEVTG 1041
Db 649 LPYLYTLNYDAHTTGAPIARPLFFSPQDPETAVYSKQFLGPGVLIISPLYNKTTSVNA 708
Qy 1042 YPPLGTWYDLQVPIEALGSLPPPPAAPREPAIHSEGWVTLPAPLDTINVHLRAGYIIP 1101
Db 709 YFPKGSWYNLNDMTM-----AVKSSGQVYTLQAPMDTINHVHCEGMILP 752
Qy 1102 LQCPGLTTTESRQCPMALAVAL-----TKGGEARGELFWDGSELEV-LERGAYTQVIFL 1155
Db 753 MQRGGMTTTVARMTPTTLIIAPLFGQSTGGKAKGHLFLDGEDVDMKIAEGKSTYVDFS 812
Qy 1156 ARNNTIVNELVRVTSIGA-----GLQLOKVTVLGVATA--POQVL--SNGVPV--SNFTY 1204
Db 813 AESDGKKVRLVSQVESGSYGLSQGWVVEKLMILGLSKSHLSQIAFQLDGKPFSSFTY 872
Qy 1205 S 1205
Db 873 S 873

Search completed: April 5, 2005, 01:05:03
Job time : 155 secs